2005

Design and development of GrainNet - universal Internet enabled software for operation and standardization of near-infrared spectrometers

Robert Dzupin
Iowa State University

Design and development of GrainNet – universal Internet enabled software for operation and standardization of near-infrared spectrometers

by

Robert Dzupin

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Agricultural Engineering

Program of Study Committee:
Charles R. Hurburgh, Major Professor
Carl J. Bern
Roger Ginder
Steven J. Hoff
Doug W. Jacobson

Iowa State University

Ames, Iowa

2005

Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Robert Dzupin

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

TABLE OF CONTENTS

ABSTRACT iv

CHAPTER 1. GENERAL INTRODUCTION 1

CHAPTER 2. IMPROVEMENT OF PREDICTION SPEED AND ACCURACY WITH INTERNET ENABLED NETWORKING SOFTWARE 10

CHAPTER 3. EVALUATION OF STANDARDIZATION ALGORITHMS FOR NEAR INFRARED SPECTROMETERS 28

CHAPTER 4. DEVELOPMENT OF A NEW NIRS CALIBRATION MODEL USING COEFFICIENTS FROM WAVELET DECOMPOSITION IN FEED –FORWARD NEURAL NETWORK ARCHITECTURE 61

CHAPTER 5. GENERAL CONCLUSIONS 77

APPENDIX A. SCREENSHOTS OF SOFTWARE USED IN THE PROJECT 79

APPENDIX B. LISTING OF THE MATLAB SCRIPT USED IN THE PROJECT 82

APPENDIX C. MATLAB CODE OUTPUTS 92

APPENDIX D. TABLES WITH RESULTS 106

ACKNOWLEDGEMENTS 117

ABSTRACT

A current trend in modern near-infrared spectroscopy is the incorporation of sophisticated mathematical algorithms into the computer instrumentation used to extract information from raw spectral data by applying complex multivariate models. To address some of the problems that near-infrared spectroscopy faces, the GrainNet software model that connects a MATLAB® computing and development environment, NIR spectrometers, and MS Server data-storage for spectral data and calibration models, was developed.

GrainNet is a client-server based Internet enabled communication and analyzing model for Near-Infrared (NIR) instruments. FOSS Infratec, Perten, and Bruins Instruments are currently three brands of the NIR instruments that have been included in the project. The performance of the implemented calibration models was evaluated. Three calibration models are implemented in the GrainNet: Partial Least Squares Regression

Artificial Neural Network

Locally Weighted Regression

The Piecewise Direct Standardization (PDS), Direct Standardization (DS), Finite Impulse Response (FIR) and Multiplicative Scatter Corrections (MSC) models were developed in the MATLAB® environment and tested for standardization transfer of the Bruins Instruments and Foss Infratec grain analyzers. A new calibration model for corn that uses feed-forward back-propagation neural networks with wavelets signal decomposition used as an input was developed.

CHAPTER 1. GENERAL INTRODUCTION

INTRODUCTION

The discovery of near-infrared energy is attributed to Herschel in 1800. The utilization of this new discovered spectrum was very limited for more than century. Only after the early 1950s, when a breakthrough in detector development occurred, has near-infrared (NIR) spectroscopy been used for chemical analysis. Credit for subsequent advances is usually given to researchers in the field of agricultural science, especially Karl Norris, who recognized the potential of this technique from the very early stages of its development.[1]

Today, NIR spectroscopy is rapidly establishing itself as a valuable technique in quantitative analysis.[2] NIR has been used successfully with many products with its biggest use in the determination of the quality traits of agricultural commodities that had never been measured before.[3]

The main difficulty for near infrared (NIR) spectroscopists over the years has been to convince the "classical" spectroscopist to accept the near-infrared measurement in the absence of a real interpretation of the spectral response.[3] It took several decades of intense argumentations before near-infrared spectroscopy became a generally accepted technique.[4]

Initial reluctance to accept NIR was caused by the fact that spectral analysis in the NIR region is not straightforward. The NIR region covers the interval between approximately 750 and 2500 nm (Figure 1). This region contains overlapped absorption bands corresponding to overtones and combinations of fundamental vibrations.[4] Their identification and assignment in the NIR region to vibrations of special molecular configurations with their

unique chemical bonds is, contrary to the situation in the mid-infrared (MIR) region, very difficult because of broad band absorbance peaks and severely overlapping vibrations.[5]

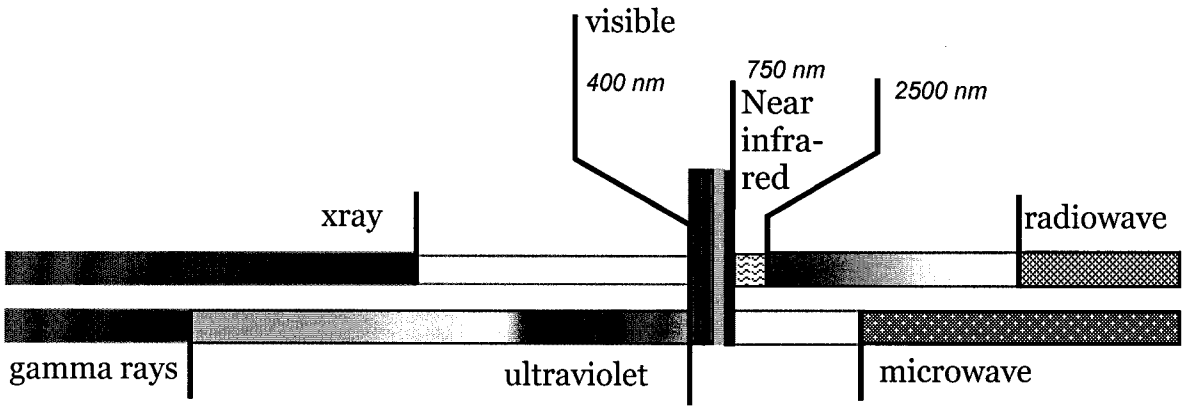

Figure 1. Spectral region of near-infrared radiation

While NIR spectra generally lack the specificity of the mid-infrared spectra, the ability to obtain quality spectra from thick samples in glass bottles and by fiber-optic probes makes NIR spectroscopy a superior technique in a number of applications. It is difficult to assign specific bands to specific chemical species, as is necessary when using the traditional univariate approach.[6] Due to the high information content of an infrared spectrum and the fact that this spectrum reflects properties of the entire molecule, development of new, more complex multivariate calibration and prediction models is necessary.[7] Acceptance of near-infrared (NIR) spectroscopic applications would be impossible without the parallel development in chemometric evaluation methods, and more specifically the advances in multivariate statistics.[4]

Chemometrics can be characterized as manipulating and investigating multiple measurements on one or many samples by applying multivariate analysis, which is essential for qualitative and quantitative analysis based on NIR spectroscopy.[8] In the NIR region,

chemometrics is used to extract useful information from NIR spectra.[9] Chemometrics has evolved rapidly over the last 15 years, largely driven by the widespread availability of powerful, inexpensive computers.[10]

In fact, most recent chemometric methods research has addressed applications of NIR spectroscopy.[11] The mathematical manipulation of experimental data is becoming a basic operation associated with NIR spectroscopy. Computerization and availability of powerful software packages is critical. Today, we are using the near-infrared measurements and software packages with chemometrics routines to analyze complex composite materials of various morphologies. The resulting spectra are often, at first sight, rather featureless, and the identification of bands and their direct use for quantitative analytical evaluation is nearly impossible.[4]

For quantitative and qualitative analyses, NIR spectroscopy needs a calibration equation. The calibration procedure involves collecting a number of samples, obtaining both reference and Near Infrared (NIR) data on each sample and developing a calibration equation that for prediction of reference results for future samples.[4] The traditional calibration technique, ordinary least-squares regression (OLS), has been replaced by more powerful methods such as Principal Component Regression (PCR), partial Least Squares Regression (PLS), and neural networks.[12]

PARTIAL LEAST SQUARES METHOD

The Partial Least Squares (PLS) regression is a technique developed and popularized in analytical science.[8] Partial least squares (PLS) together with principal component

regression (PCR) are the most widely used multivariate calibration methods in chemometrics. Both of these methods make use of the inverse calibration approach.[6]

PLS regression is a multivariate calibration method that includes the dependent (e.g., protein concentration) variable in the data compression and decompression operations. PLS is designed to deal with highly correlated data, such as near-infrared spectra. The strategy is not to select a subset of less correlated features but rather to consider highly correlated features as multiple measurements that increase the stability of the model.[12]

The criterion mostly applied in PLS is maximum covariance between latent variables and tested property. PLS is a linear method and therefore, the final latent variable that is used to predict the property is a linear combination of predicted features.[12]

The regression equations for both PLS and PCR, based on the centered (adjusted to a zero mean) y and **x**, can be written as:

Equation 1

$$y = q_0 + \sum_{a=1}^{A} q_a \left(\sum_{k=1}^{K} w_{ka} x_k \right) + f = q_0 + \sum_{a=1}^{A} q_a t_a + f$$

where:

y	is the output variable
f	is a random variable
w_{ka}	are functions of the loadings and loading weights
t_a	is the latent variable
q_a	is the regression coefficient of y on the latent variable
x_k	spectral data in each of K wavelengths

ARTIFICIAL NEURAL NETWORKS

An alternative to linear PLS and PCR methods is the Artificial Neural Networks (ANN) calibration model. The acronym ANN originates from Artificial Intelligence (AI) research and was used to model how networks of interconnected neurons in the human brain produce intelligent behavior.[13]

ANN for calibration of the near-infrared spectra in agriculture has feed-forward architecture. Feed-forward networks have one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons. The nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output variables.

The feed-forward network structure corresponds to a regression equation[13] of the form:

Equation 2

$$y = h\Big[\sum_{a=1}^{A} q_a g_a\Big(\sum_{k=1}^{K} w_{ka} x_k + \alpha_{a1}\Big) + \alpha_2\Big] + f$$

where:

y is the output variable

f is a random variable

g, h are specified functions

w_{ka} are weights that each input element that must be multiplied

A is number of nodes

K is number of elements

q_a, α_{a1}, α_2 are parameters to be estimated from the data

As seen from the equation, an artificial feed forward neural network is simply a non-linear model for the relationship between y and all the x-variables. The regression equations for both PLS and ANN (Equation 1 and Equation 2), apart from the non-linear g_a and h in the feed-forward network, are identical. The equation for PLS (Equation 1) is a special case of an ANN equation (Equation 2) with linear g_a and h. Missing constants α_{a1}, α_2 in the PLS equation (Equation 1) are due to the centering of y and x in the PLS model.[13]

The main difference between PLS and ANN is how the weights w_{ka} in the PLS and ANN equations are determined. The PLS estimates those parameters by maximizing the covariance between y and linear functions of x, while the ANN regression used for NIR calibration estimates those parameters without restrictions by using back-propagation. The ANN models are therefore, more prone to overfitting than are PLS models.

To reduce overfitting in ANN models, data compression of input variables used in the model, is often implemented. A very popular method used for compression of the input data is the linear compression method used in PLS, Principal Component Analysis (PCA). When relationships between input and output variables are nonlinear, linear PCA compression is undesirable. A new compression method capable of preserving non-linearities is the ANN regression that uses wavelet decomposition for inputs. This method was developed and tested in the third part of this research.

DISSERTATION ORGANIZATION

This dissertation is written in the alternate dissertation format with an introduction followed by three papers formatted for submission to the Journal of NIR Spectroscopy:

1. Improvement of Prediction Speed and Accuracy with Internet Enabled Networking Software.
2. Evaluation of Standardization Algorithms for Near-Infrared Spectrometers.
3. Development of a new NIRS Calibration Using Coefficients from Wavelet Decomposition in Feed-Forward Neural Network Architecture.

A general conclusion and recommendations follows the papers.

REFERENCES

---

[1] W. F. McClure, "Near Infrared Instrumentation", Near-Infrared Technology in the Agricultural and Food Industries, Edited by Phil Williams and Karl Norris, ©2001 by the American Association of Cereal Cheminsts, Inc, p. 109

[2] Sumio Kawano "Application to Agricultural Products and Foodstuffs", "Near-Infrared Spectroscopy. Principles, Instruments, Applications", Copyright © WILEY-VCH Verlach GmbH, 2002, p.289

[3] Franklin E. Barton, "Progress in near infrared spectroscopy the people, the instrumentation, the applications", Near Infrared Spectroscopy: Proceeding of the 11th International Conference, Edited by A.M.C Davies and A. Garrido-Varo, p.16

[4] H.W. Siesler, S.Kawata, H.M.Heise "Origin of Near-Infrared Absorption Bands", "Near-Infrared Spectroscopy. Principles, Instruments, Applications", Copyright © WILEY-VCH Verlach GmbH, 2002

[5] Hais F. Parreira, Marcia M. C. Ferreira, Henrique J. Sales, Wanderson B. Almeida, "Quantitative Determination of Epoxidized Oil Using Near-Infrared Spectroscopy and Multivariate Calibration", Applied Spectroscopy, © 2002 Society for Applied Spectroscopy, Volume 56, Number 12, 2002, p. 1607

[6] Kenneth R. Beebe, Randy J. Pell, May Beth Seasholtz, "Chemometrics A Practical Guide", A Wiley-Interscience Publication,1998

[7] Paul Selzer, "Correlation between Chemical Structure and Infrared Spectra", "Handbook of Chemoinformatics – From Data to Knowledge" , Edited by Johann Gasteiger, Copyright © WILEY-VCH Verlach GmbH, 2003, p.1349

[8] M.J. Adams, "Chemometrics in Analytical Spectroscopy 2nd Edition, RSC Analytical Spectroscopy Monographs, Published by Royal Society of Chemistry, 2004

[9] Koichi Myrayama, Bogslawa Czarnik-Matusewicz, Young Wu, Rouyama Tsenkova, Yukihiro Ozaki," Comparison between Conventional Spectral Analysis Methods, Chemometrics, and Two-Dimensional Correlation Spectroscopy in the Analysis of Near-Infrared Spectra of Protein", Applied Spectroscopy, Volume54, Number 7, 2000, p.978

[10] Richard.Kramer, "Chemometric Techniques for Quantitative analysis", Copyright © 1998 by Marcel Dekker, Inc., p.1

[11] Stephen R. Lowry, Jim Hyatt, and William J. McCarthy, "Determination of Wavelength Accuracy in the Near-Infrared Spectral Region Based on NIST's Infrared Transmission

Wavelength StandardSRM 1921", Applied Spectroscopy, © 2000 Society for Applied Spectroscopy, Volume 54, Number 3, 2000, p. 450

[12] Kurt Varmuza, "Multivariate Data Analysis in Chemistry", "Handbook of Chemoinformatics – From Data to Knowledge" , Edited by Johann Gasteiger, Copyright © WILEY-VCH Verlach GmbH, 2003

[13] Tormod Naes, Thomas Isakson, Tom Fearn, Tony Davies, "A user-friendly guide to Multivariate Calibration and Classification", © 2002 NIR Publications

CHAPTER 2. IMPROVEMENT OF PREDICTION SPEED AND ACCURACY WITH GRAINNET- INTERNET ENABLED NETWORKING SOFTWARE

A paper to be submitted in the Journal of NIR Spectroscopy

Robert Dzupin, Charles. R. Hurburgh and Sylvie. A. Roussel

Agricultural and Biosystems Engineering Department, Iowa State University,
1551 Food Sciences Building, Ames, IA 500 14, USA

ABSTRACT

The ISU Grain Quality Laboratory has been creating calibrations for near-infrared (NIR) analyzers. Through this process, very large databases, containing information on thousands of samples, have been collected. Very large data sets and fast computers allow the use of mathematical methods and multiple models not supported by internal instrument software. A software solution (GrainNet) was designed to:

- Implement a universal Internet-enabled communication and analysis model for NIR instruments of any brand.
- Create a model for handling of data through Internet-capable storage to provide immediate analytical results for unknown samples and store spectra in a central database.
- Develop a scalable object-based system of implementation for data processing and analysis.
- Implement and compare multiple mathematical algorithms in real time.

The software links multivariate instruments with high capacity numerical software (MATLAB™) for central server processing over the Internet. The combination greatly enhances measurement capabilities and automates data inventory management. NIRGrainNet was tested in the fall of 2001, using corn and soybean samples on which moisture and protein was being measured with 3 Foss Infratec 1229/1241 analyzers (FOSS, www.foss.dk).

Spectral data and predictions on three models (partial least squares, locally weighted regression and artificial neural network) were captured, reported in real time and compared. As expected, the nonlinear models were more accurate than the PLS models, but the best accuracy was obtained by either selecting the best model for each sample/constituent situation or by averaging the results of the three models. Thus, real time access to rapid computing can improve accuracy by merging prediction outputs of several cutting-edge chemometrics models as well as facilitate operations of instrument network management.

INTRODUCTION

Near-infrared (NIR) instruments are popular for the prediction of chemical composition and biological properties of food and agricultural material. In the agricultural and food industries, NIR instruments are primarily used for the detection of C-H, N-H and O-H bonds, which relate to concentration of oil, protein and moisture. The advantages of using NIR instruments are that near-infrared spectroscopy is an unusually fast technique compared to other analytical techniques (often taking less than 1 minute), it is nondestructive, and minimal sample preparation is required. The standard use of NIR spectroscopic data relies on

the development of multivariate calibrations. This has been a serious restriction of NIR spectroscopy applications because of the high cost of calibration development. Typically 300 of more samples with reference data, up to a thousand or more for ANN are needed to develop calibration models[1,2,3].

NIR spectroscopic data are used to predict analyte values and to construct a calibration model in the form of a regression equation. This equation can then be used to predict unknown samples from NIR measurements. The equation is usually obtained by a partial least-squares regression (PLS),[3] a well-established multivariate linear method.

However, this calibration technique cannot model non-linearities. A major concern when building a model based on measurements coming from a single master NIR instrument is the transferability to the other units. Calibration transfer inherently introduces non-linearities because instrumental variations are not necessarily linear. Non-linear calibration methods could improve the accuracy of prediction models as well as their inter-instrument transferability.

Local modeling reduces the need for expensive calibration derivation and update.[4] Instead of using a regression equation to summarize the database, the complete database is employed. Alternatively, an artificial neural network (ANN) can be used. Both calibration approaches depend on the accumulation of a very large database, with each item possessing full spectra and analytical data.[5]

Nonlinear and large database local models can be implemented over the Internet. Software was designed to provide environment for database analysis calculations in the real time. Beside internet connectivity, the solution assumed that the NIR spectrometer will provide a communication interface to send measured optical data to a personal computer. In

the current setting the RS 232 interface (serial port) was used to establish a link between instrument and personal computer. The prediction is done by a remote server[6]. The local PC only provides communication and data management. Centralized calculation of this solution also allows simultaneous prediction of the same constituent by several models. It is likely that individual samples are better predicted by one model over others. If a model selection routine can be developed, overall accuracy would be improved by matching samples to models.

The objective of this paper is to evaluate performance of the real-time centralized system for handling of data over Internet developed in Grain Quality Laboratory and to explore possibility of improvements of accuracy by merging prediction outputs of several chemometrics models implemented in the system.

MATERIALS AND METHODS

DESCRIPTION OF THE SOFTWARE

The main concept is to link NIR spectrometers and a commercially available database management system (SQL Server™) with flexible, high capacity numerical software (MATLAB™). In MATLAB™ (The MathWorks Inc., www.mathworks.com), additional calculations and data management routines can be implemented.

The software (Fig. 1) has three components:

1. Client computer – used to retrieve optical data from NIR spectrometer, send them over Internet to central database SQL server using modem, DSL or T1 connection. The client

computer requirements are; MS Windows9X, ME, 2000 or XP operating system and a PC that can support the selected operating system. In our testing environment, IBM PC computers with 66 MHz processor speed running Windows 95 proved to be sufficient.

2. A computer running model calculations in Matlab™ – Personal computer with fast processor (Pentium III or IV) used to process linear, non-linear or database models and calculate predictions. Matlab server is connected to central database SQL server. If real-time processing is required, connection speed requirements are higher than for client computer. (T1, T3 or LAN)

Matlab™ computer requirements: The computer running MATLAB models determines if system can be used in real time, therefore only Pentium III processor with 800 MHz processor or faster have been used in the software system. Because software is using MS Windows specific API calls, only MS Windows9X, ME, 2000 or XP are supported. Windows 2000 and XP are recommended. During our laboratory testing, Windows9X was an unstable platform for running MATLAB™ routines over extended period of time.

3. SQL server – The database server that stores optical data, sample identification data, and calculated predictions from Matlab™ model. The requirements are: A MS SQL 7.0 or 2000 Database server requirement for small systems (less than 50 concurrent connections) is similar to Matlab™ computer requirements. Database operations are characterized as input/output very intensive. Therefore SCSI hard drives, preferably using RAID arrays[7], are recommended.

Optical data retrieved from NIR instruments are accompanied by Instrument ID, Time, Computer ID, User Name, and by data manually entered by the operator (Sample ID,

Variety, etc.). Therefore, each set of optical data in the SQL Server© database can be uniquely identified, as required for instrument network management.

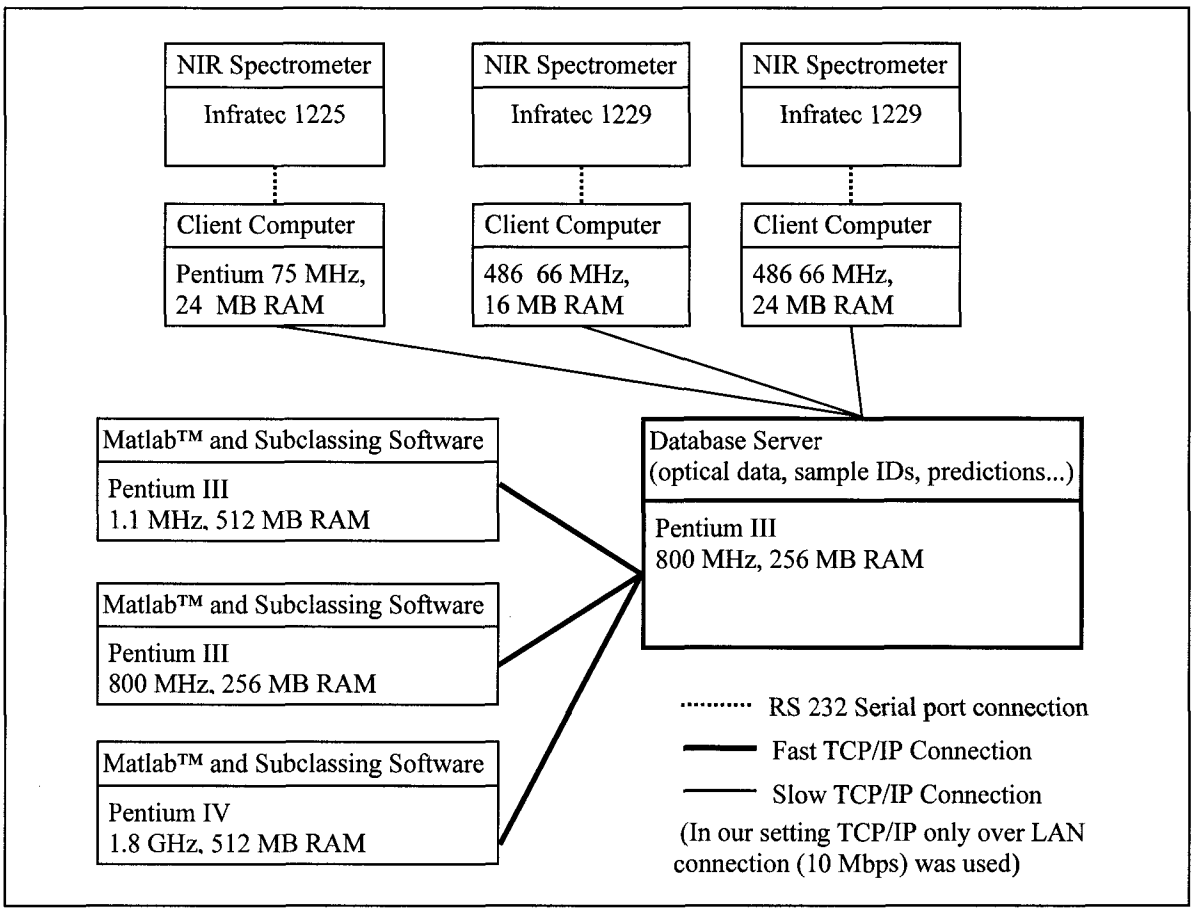

Figure 1. GrainNet software Setup (as of July 2000)

NEAR INFRARED SPECTROMETERS

Three near infrared spectrometers Foss/Tecator Infratec instruments (1225-Infratec serial 0065 and two 1229-Infratec serial 553075 and 243108 were used to collect transmission spectra of whole corn samples(Fig. 2). Spectrometers provide 100-wavelength spectra in the 850-1050nm range, with 2-nm resolution. The calibration database contains measurements provided by 3 Master instruments 1225-Infratec #0065 with a cuvette sample

presentation, 1229-Infratec #553075 and 1241-Infratec #0350 with a flow presentation. Five constituents (moisture, protein, oil, starch, and density) were reported.

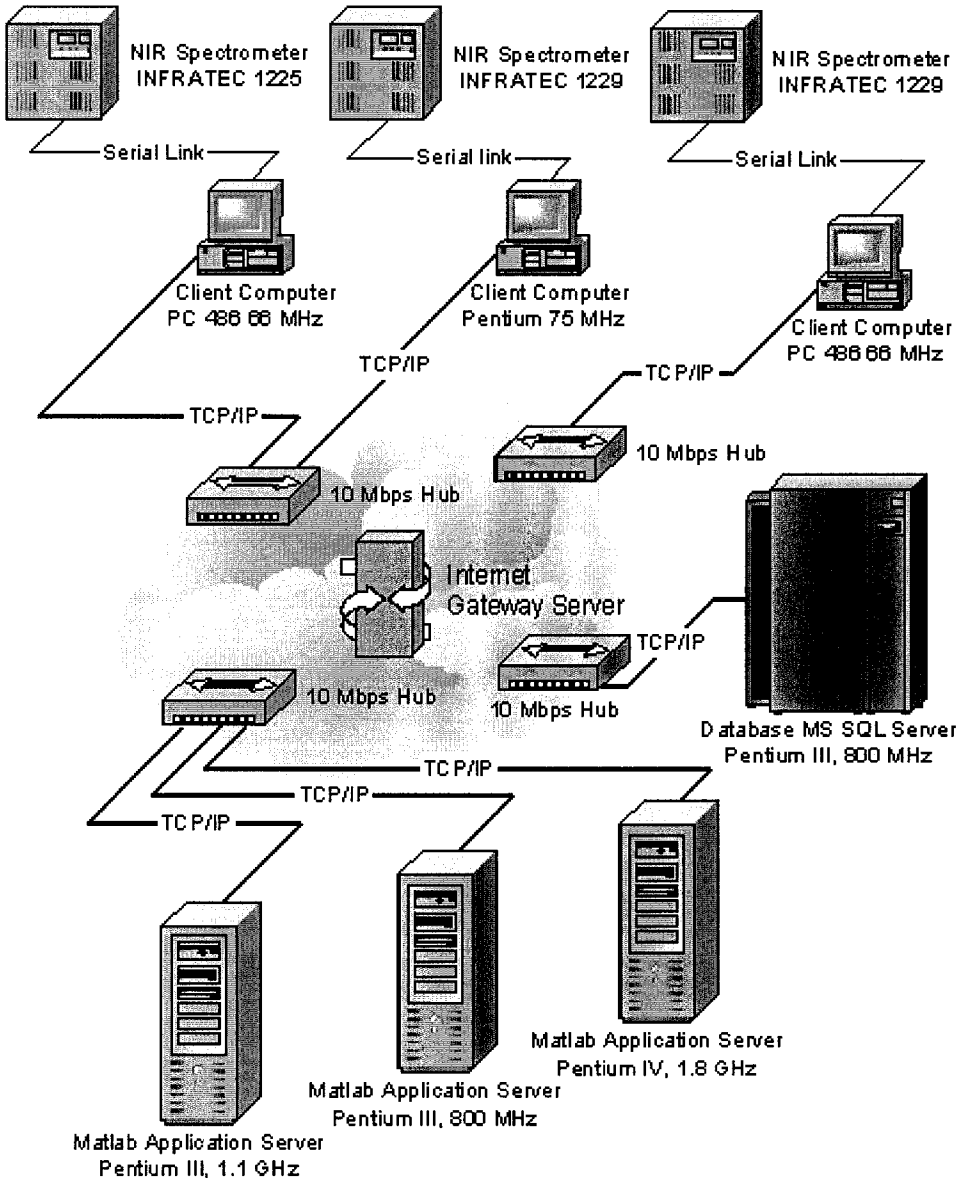

Figure 2. Grain-Net hardware configuration (as of July 2000)

PROCESSING ALGORITHMS

Three processing algorithms were implemented in this test:

- a linear regression model (Partial Least-Squares Regression: PLS),

- a local regression model (Locally Weighted Regression: LWR) and
- a non-linear model (Artificial Neural Networks: ANN).

There is no restriction on the number of models or processing algorithms that could be used.

Partial Least Squares

Partial Least Squares Regression (PLS) is a well documented multivariate linear model that is well documented and commonly applied in NIR.[8] In this study, PLS is the reference model for comparison. All the data were mean-centered and the number of latent variables was tuned (lvs $\leq$ 15). In our model, 13 latent variables were used.
To reduce the number of wavelengths and increase the robustness and transferability, the Standard Normal Variate (SNV)[9] pre-processing technique was applied.

Locally Weighted Regression

The Locally Weighted Regression (LWR)[10] builds local linear regressions that enable the model to fit non-linearities. For each sample, its neighborhood is determined by the Mahalanobis distance computed on the first principal components issued from x-values (spectra) and the Euclidean y-distances. Since the y-values of the samples to be predicted are unknown, the distance and the neighborhood are computed iteratively. The neighborhood size as well as the weighting given to the distance in y (alpha) must also be tuned carefully.[11] In the GrainNet software implementation, the lwrxy function from The MATLAB™ PLS toolbox (Eigenvector Research Inc., www.eigenvector.com) was used as the LWR model[10].

Input parameters used:

lvs - the number principal components used to model the independent variables
npts - the number of points defined as local
alpha - the weighting given to the distance in y
iter - the number of iterations to use

These were determined in previous study by Roussel et al[3].

Table 1: Parameters for locally weighted regression

CORN	Moisture	Protein	Oil	Starch	Density
Lvs	14	15	15	15	15
npts	300	500	300	250	300
alpha	0.9	0.9	0.9	0.9	0.8
iter	2	2	2	2	2

Artificial Neural Network

Artificial Neural Networks (ANNs) are able to fit non-linear relationships between multivariate x and y-values. In this study, supervised 3-layer feed-forward neural networks are trained with dynamic learning using error-gradient back-propagation algorithms.[12] The inputs (and the outputs) are scaled between –1 and +1 to fit to the range of the hyperbolic tangent activation functions.

The master database was used to train the ANN, with no early stopping method (to prevent stop any training too early because the error descent is not monotonous). Instead, the number of epochs was tuned. In our model, the neural network contained 30 inputs, 10 hidden layers, and 2500 epochs were used[13].

CALIBRATION DATABASES

The database contained 6442 corn samples : 2762 from unit serial 0065, 2823 from unit serial 553075, and 857 from unit serial 0350.

Database cleaning outlier were removed with PCA (spectral outliers) and prediction residuals (chemistry value outliers) for every constituent.

Table 2: Corn calibration database and models (January 2001)

CORN	Moisture (as-is)	Protein (as-is)	Oil (as-is)	Starch (as-is)	Density (g/cm^3)
Initial database	5782	2138	2137	2127	1925
PCA outliers	3	1	1	1	1
Residual outliers	9	21	12	12	64
Final database	5782	2116	2124	2062	1857
Calibration set	4625	1693	1699	1649	1485
Test set	1157	423	425	413	372
SEP for LWR model	0.32% pts	0.33% pts	0.30% pts	0.70% pts	1.63% pts
# factors for LWR	14	15	15	15	15
# neighbors for LWR	300	500	300	250	300
Alpha used in LWR	0.9	0.9	0.9	0.9	0.8
SEP for NN model	0.31%	0.28%	N/A*	N/A*	N/A*
SEP for PLS Model	0.41%	0.34%	N/A*	N/A*	N/A*

*Model was not developed

MODEL COMPARISON

Set of 30 samples with wet chemistry references provided by Woodson–Tenent Laboratories, Inc. (Des Moines, IA) was used to explore possibilities for improvement in robustness and precision of the models. These samples had replicated chemistry values and were laboratory transfer standards and were not part of calibration databases. The corrected standard error of prediction (SEP corrected) was calculated.

Bias corrected standard error of prediction was calculated by the equation:

$$SEP(corrected) = \sqrt{\frac{\Sigma(y-x)^2 - (\Sigma(y-x))^2 / n}{n-1}}$$

where: y is the result from the chemical analysis

x is the result predicted from NIR measurements

n is the number of samples in the validation set

SEP(corrected) was calculated for PLS, ANN and LWR models. SEP was also calculated for the average of prediction differences of all three models. An optimal SEP was manually calculated. From the model that was closest to the chemical analysis result for each sample individually.

SYSTEM PERFORMANCE

To estimate the number of instruments that could supported by GrainNet software, the throughput of the system (number of processed samples per minute) was calculated:

$$Throughput = \frac{60}{t} t_s$$

where:

t = t1+t2+t3+t4+t2

t1 – time (in seconds), necessary to retrieve data from SQL Server™ database to computer running Matlab™

t2 – network delay (in seconds) between SQL Server™ database and Matlab™ computer

t3 – time (in seconds), needed to processing data in to Matlab™ environment

t4 – time (in seconds), necessary to update SQL Server™ database with output from Matlab™

ts – time (in seconds) to measure one sample on the NIR spectrometer (load, measure and unload sample from spectrometer)

RESULTS

MODEL PERFORMANCE EVALUATION

To compare the performance of the PLS, ANN and LWR models, SEP(Corrected) was calculated (Table 4). As expected, the ANN and LWR models were more accurate than the PLS model. SEP was also calculated for the average prediction of all models. LWR was the model with lowest SEP when processing optical data collected from 1225-Infratec #0065 spectrometer. It is also the only unit using the cuvette configuration.

The ANN method had lowest SEP for 1229-Infratec 553075 and 243108 spectrometers, but in the same time the SEP for 1225-Infratec 0065 using ANN was the

highest of all three models even though this instrument was in the training database. The SEP for averaged prediction differences was more consistent across units. To represent what might be ideally achieved with model selection, the optimal model concept was introduced. In the optimal method, the prediction closest to the reference value is manually selected from the pool of models.

Table 3. Corrected Standard Error of Prediction for corn protein (January 2001)

SEP	PLS model	LWR model	ANN model	Model with Averages	Optimal model
Spectrometer 0065	0.31	**0.28**	0.31	0.29	0.24
Spectrometer 553075	0.28	0.26	**0.24**	0.25	0.22
Spectrometer 243108*	0.30	0.29	**0.27**	0.28	0.22
Average	0.30	0.28	0.27	0.27	0.23

Number of samples: 30
*Not in the calibration pool

Table 4. Corrected Standard Error of Prediction for corn protein (July 2004)

SEP	PLS model	LWR model	ANN model	Model with Averages	Optimal model
Average 3 Spectrometers*	0.33	0.34	0.36	0.32	0.24

Number of samples: 135
Updated PLS and ANN calibration model used (June 2003)
* Spectrometers used: 0065, 553075, 243108

SYSTEM PERFORMANCE EVALUATION

Throughput of the models is reported in Table 3. The first line shows throughput when all 3 processing algorithms were used. The second line of the table shows throughput of the system, using only one database processing algorithm (LWR) to calculate five constituents (moisture, protein, oil, starch, and density). The database throughput was also measured. Database throughput is the number of samples that can be processed by computer used in the system if no model calculation is performed. Database throughput accounts for network delays between the database and servers with Matlab™ routines. Because the computers are using same network connection to the database server, database throughput is same for all three computers.

The data in Table 5 can used to estimate the required number of computers for processing selected calculations in Matlab™, in for real time support of the NIR spectrometers. The assumption is that the new optical data are sent from the NIR spectrometer once per minute. If this time is different, the estimated number of processed samples needs to be multiplied by the appropriate ratio. For example, if the processing time for one sample is three minutes, number of users that computer can handle would be three times higher. First line of the table shows group of models integrated into GrainNet software in July, 1999. Second line of the table shows performance characteristics of the model that was integrated into GrainNet year later in July, 2000.

Table 5. Throughput of implemented models

Model	Pentium 0.8 GHz (Processed samples/minute)	Pentium 1.1 GHz (Processed samples/minute)	Pentium 1.8 GHz (Processed samples/minute)	Database throughput (Processed samples/minute)
PLS ANN LWR (2 constituents)	11.6	13.8	16.2	35.0
LWR corn (5 constituents)	16.1	16.4	22.1	35.0

DISCUSSION

An "ultimate" system where calibration is based on samples supplied by diverse clients to a host laboratory, and is used to predict results upon receipt of spectra by e-mail, using the local or ANN models, was proposed by Phil Williams[3]. GrainNet software is extending the idea of the "ultimate" system to real-time and opens the possibility of improving accuracy of prediction by center averaging the results of several models or choosing models based on sample properties. Creation of selection algorithms is the subject of other studies.

Because the NIR instruments collect raw optical data, GrainNet software is not limited to any particular NIR instrument manufacturer. The only implementation requirement of the instrument is the capability of the spectrometer to send raw optical data to a standard communication port. (RS 232, USB, etc.)

The software requires a fast network connection between the database server and computers that process the models in Matlab™. A fast network connection is especially necessary if several computers are used to calculate prediction. For example, to predict 5

constituents using LWR, we can process 54 samples per minute with three computers instead of 16 or 22 samples per minute if only one computer is used.

Data in Table 5 suggests that real-time access to rapid computing can improve accuracy by merging or selecting among prediction outputs of several chemometrics models. Using the optimal model to estimate the potential improvement beyond the PLS, LWR, or ANN models, the accuracy of all three models can be improved. The accuracy of the PLS model was improved by 23 percent. The accuracy of the LWR model was improved by 18 percent and the accuracy of the ANN model was improved by 15 percent. The possibility of improvement of model accuracy was confirmed by measuring corn protein SEP on a validation set of 135 samples predicted by using updated ANN and PLS calibration models developed in Grain Quality Laboratory in 2003 (Table 4). The accuracy of the PLS model on the validation set of 135 samples was improved by 27 percent. Similarly, the accuracy of the LWR model was improved by 29 percent and the accuracy of the ANN model was improved by 33 percent.

REFERENCES

[1] Pierre Dardenne, George Sinnaeve and Vincent Baeten, Multivariate calibration and chemometrics for near infrared spectroscopy: which method?, J. Near Infrared Spectroscopy, 8, 229-237 (2000)

[2] Stephen R. Delwiche, Robert A. Graybosch, and C. James Peterson, Predictiong Protein Composition, Biochemical Properties, and Dough-Handling Properties of Hard Red Winter Wheat Flour by Near-Infrared Reflectance, Cereal Chemistry, Volume 75, Number 4 (1998)

[3] Sylvie A. Roussel, Glen R. Rippke, Charles R. Hurburgh, Jr., Accuracy and robustness comparison of different processing algorithms for grain quality assessment, Proceedings Pittcon 2000., www.pittcon.org (2000)

[4] Tony Davies, Some variations on a 'Local' theme, Spectroscopy Europe, July/August 1999, Norwich Near Infrared Consultancy, 75 Intwood Road, Cringleford, Norwich NR4 6AA, UK. E-mail: td1@nir-pub.demon.co.uk (1999)

[5] Phil Williams and Karl Norris, Near-Infrared Technology in the Agricultural and Food Industries, American Association of Cereal Chemists, Inc. St. Paul, Minnesota, USA, p. 145 (2001)

[6] Robert Dzupin, "Client-server base Internet communication and analyzing software for NIR instruments", Iowa State University, Ames, Iowa (2002)

[7] Brad M. McGehee, Rob Kraft, MatthewShepker, Eric Wilson, Simon Gallagher and Tibor Karaszi, "Practical Microsoft SQL Server 7", © 1999 Que® Corporation

[8] Geladi P, Kowalski B.,Partial Least Squares regression: a tutorial, Anal. Chim. Acta, 185, pp1-17 (1986)

[9] Barnes RJ, Dhanoa MS, Lister SJ,Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra, Appl. Spectrosc., 43(5), pp772-777 (1989)

[10] Wang Z, Isaksson T, Kowalski BR., New Approach for Distance Measurement in Locally Weighted Regression, Anal. Chem., 66, pp294-260 (1994)

[11] Sylvie A. Roussel, Glen R. Rippke, Charles R. Hurburgh, Jr., Accuracy and transferability comparison of various multivariate processing algorithms for corn quality assessment, The 10th International Diffuse Reflectance Conference (2000)

[12] Rumelhart D E, Learning and generalization, In Proceedings of IEEE International Conference on Neural Networks, San Diego, CA, USA (1988)

[13] Sylvie A. Roussel, Glen R. Rippke, Charles R. Hurburgh, Comparison of Global Linear, Local Linear and Non-Linear Models for Grain Quality Assessment, Proceedings Pittcon 2003., www.pittcon.org (2003)

CHAPTER 3. EVALUATION OF STANDARDIZATION ALGORITHMS FOR NEAR INFRARED SPECTROMETERS

A paper to be submitted in the Journal of NIR Spectroscopy

Robert Dzupin and Charles. R. Hurburgh

Agricultural and Biosystems Engineering Department, Iowa State University,
1551 Food Sciences Building, Ames, IA 500 14, USA

ABSTRACT

Development of multivariate calibration for quantitative regression models for agricultural products is a very expensive and time consuming process. Calibration developed on one instrument will often fail if loaded without adjustments into another instrument even if they are same brand. In this research, several different standardization approaches were tested and compared with two brands of instruments of near-infrared transmittance instruments.

Standardization transfer was tested as a part of the development of the custom GrainNet networking software. GrainNet is the software designed to connect different instruments from different brands into one network and allows the creation of a system of networked instruments that share calibration models developed in MATLAB®.

Optical standardization methods such as Direct Standardization (DS), Piecewise Direct Standardization (PDS), and standardization approach that does not need standards – Finite Impulse Response (FIR) were tested.

The data preprocessing, Multiplicative Scatter Correction (MSC) was also tested for its ability to remove differences between instruments caused by the light scatter effect. An advantage of the MSC method is that it does not require standardization samples, but a disadvantage is that this method is designed to remove mostly inter-instrument variability and is commonly listed as a pre-treatment and not standardization method.

INTRODUCTION

The development of NIRS calibrations for agricultural commodities is very costly. The main cost comes from reference method analyses. Reference methods used to develop calibration equations are slow, time consuming, and require expensive devices and reagents.[1] Powerful statistical methods such as Principal Component Regression (PCR), Partial Least Squares (PLS) regression, and nonlinear neural networks require hundreds or thousands of samples to establish a relationship between the matrix of spectral data and the vector of predicted variables.[2]

Multivariate calibration for near infrared spectroscopy utilizes the multivariate advantage. Signal averaging, where a standard deviation of a measurement is reduced by a factor of $\sqrt{n}$ when the average of n measurements is used. This allows the use of many nearly redundant measurements to construct a more precise calibration model.[3]

The result is that calibrations require large databases with very often more than 1000 samples per commodity. The development of a calibration model cannot be reproduced separately for each instrument copy. The calibration model must be usable in many instruments of the same make and model.

Generally, the responses from two spectrometers of the same model will not be the same. Differences can be traced to major instrumental effects resulting from changes in the wavelength scale and the ordinate axis, detector nonlinearity, differences in the accessory optics that influence spectral intensities and other factors.[1,4] Among the most significant engineering factors influencing instrument spectral intensity differences are preamplifier and amplifier gain settings and bandwidth.[5] Because of these differences, some form of adjustment between different instruments, even between instruments of the same brand and model, is necessary.

Currently, the most popular method for calibration transfer in agriculture is post-regression slope and bias correction. A disadvantage of this method is that it can only be used for simple instrument variability that causes a change in bias or slope. Slope and bias should be applied only in cases when a linear relationship (Constant x axis shift for example) exists and will remain stable between both instruments.[6] With large investments in calibration development, it is important that the same calibration model can be transferred to any instrument and is not limited to instruments with a linear relationship of differences between their spectra.

In order to keep discrepancies between spectrometers as low as possible, usually only instruments of the same type are used in a network.[7] With a growing number of near-infrared

instruments introduced to the market, reducing the differences between instruments from different brands and avoiding costly recalibration is critical.

An alternative approach is to build large databases that are robust enough to simplify or fully eliminate the need for standardization. This approach was also tested by using a Locally Weighted Regression (LWR) model implemented in GrainNet software[8]. Similar approaches are currently being explored by several authors.[7,9] GrainNet software was designed and developed in the Grain Quality Laboratory as a communication and analysis model for NIR instruments (Figure 1).

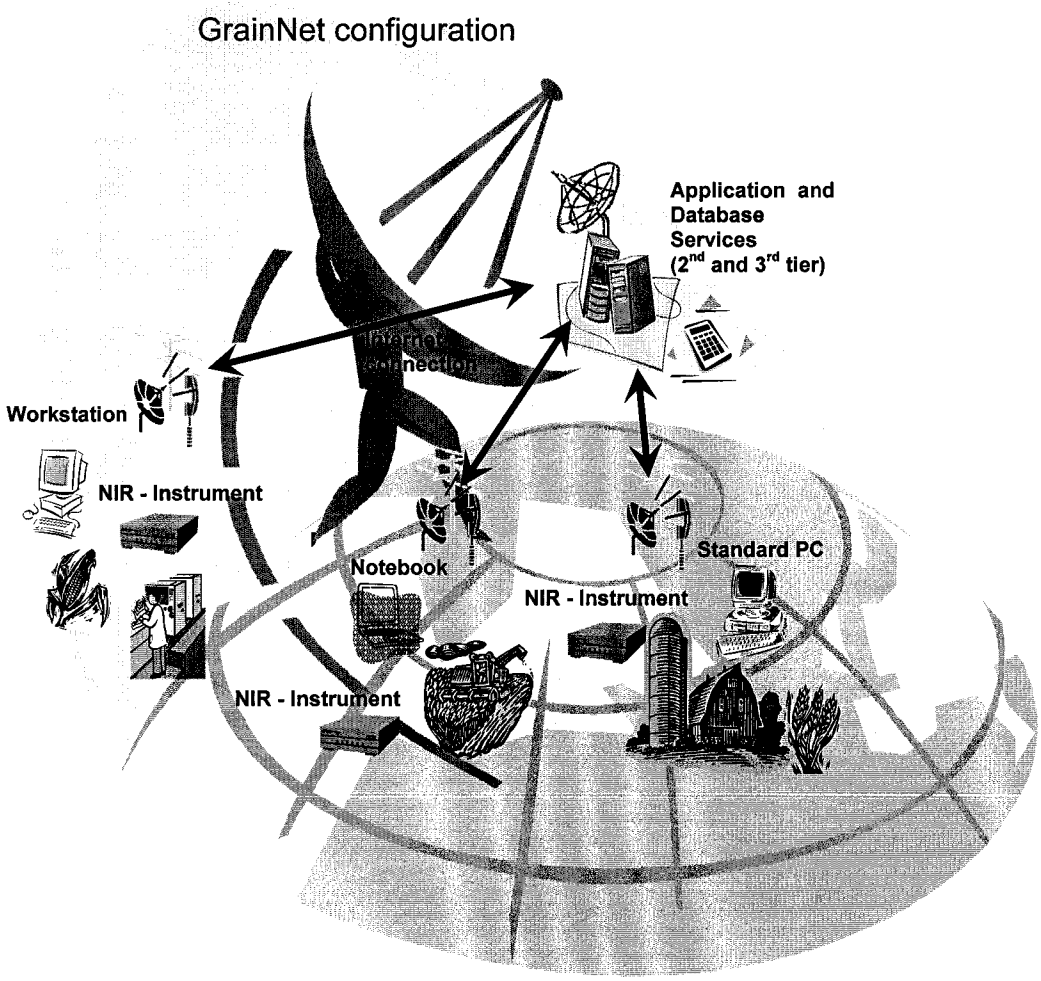

Figure 1. GrainNet communication software

Instruments of different brands connected through GrainNet software to a library of chemometrics routines are an ideal environment to develop and compare different standardization routines. Several optical standardization models were evaluated using Foss Infratec (FOSS, www.foss.dk) and Bruins Omega transmission (Bruins Instruments, www.bruins.de) analyzers. Optical standardization typically uses a set of samples that are used to develop a mathematical relationship between spectra of two different instruments. The developed mathematical model is then used to remove instrument specific differences between spectra.

Direct standardization and Piecewise Direct Standardization are two optical standardization methods with standardization samples that were tested in this project. An optical standardization method that does not require a set of standardization samples, the Finite Impulse Response Filters (FIR) method, was also developed and evaluated. The last method used in this project to minimize differences between spectra of different instruments was the Multiplicative Scatter Correction preprocessing method.

OBJECTIVES

- Using Foss Infratec and Bruins Omeg transmittance analyzers, implement and evaluate optical standardization models, Direct Standardization and Piecewise Direct Standardization for Bruins Instruments analyzers
- Implement and evaluate standardization methods that do not require standardization samples: Finite Impulse Response (FIR) standardization and mathematical pre-treatment Multiplicative Scatter Correction (MSC)

- Explore the possibility of calibration transfer between Bruins Instruments s and Foss Infratecs instruments

MATERIAL AND METHODS

SAMPLES AND ORGANIZATION

This study was concluded with 520 corn samples from 1999 to 2003 crop years. All samples had wet chemistry references provided by Woodson–Tenent Laboratories, Inc. (Des Moines, IA). The samples were divided into three groups (Table 1). The first group, the calibration set, was used to create calibration models.

Table 1. Calibration set description

Calibration set			
Property	Moisture (%)	Protein (%)	Oil (%)
Range	8.1 – 23.0	5.5 – 16.4	2.24 – 13.2
Moisture Base	As is	Direct 15%	Direct 15%
N	480	480	480
Standardization set			
Property	Moisture (%)	Protein (%)	Oil (%)
Range	8.2 – 16.3	6.7 – 12.1	3.2 – 13.1
Moisture Base	As is	Direct 15%	Direct 15%
N	20	20	20
Validation set			
Property	Moisture (%)	Protein (%)	Oil (%)
Range	10.0 – 15.8	5.5 – 12.5	3.3 – 7.0
Moisture Base	As is	Direct 15%	Direct 15%
N	20	20	20

The second group was used for development of the standardization model, and was selected by using the multivariate leverage sample selection method described later. The third group was used as a validation set (Table 2). The validation set consisted of samples routinely used in the Iowa State University (ISU) Grain Quality Laboratory for standardizing (by slope and bias) instruments. This set was selected based on variation of reference variables to cover a full range of reference values.

Table 2. Samples set description

	Calibration Set	Standardization Set	Validation Set
1996	0	0	1
1998	0	0	2
1999	145	3	2
2000	137	5	1
2001	98	4	5
2002	25	4	5
2003	75	4	4
SUM	480	20	20

INSTRUMENTS

Bruins Instruments OmegAnalyzer G (Serial 6110) and Foss Infratec 1241 Grain Analyzer (Serial 0350) instruments were used to collect spectra from all measured sample sets (Table 3). The remaining instruments were used to measure only the samples from the validation set. All three data sets were measured over the course of two weeks by one operator. The third set was measured on all tested instruments in one day.

Table 3. Description of used instruments

Bruins Instruments				
Instrument Type	Omeg AnalyzerG	AgriCheck	Omeg AnalyzerG	Omeg AnalyzerG
Serial Number	6110	31002	6118	6175
Technology	Scanning monochromator working in transmittance mode			
Spectral range	730 – 1100 nm	730 – 1100 nm	730 – 1100 nm	730 – 1100 nm
Sampling Interval	0.5 nm	0.5 nm	0.5 nm	0.5 nm
Instrument Role	Master	Slave	Slave	Slave
Number of Subsamples	16	10	16	16
Used for Calibration	YES	NO	NO	NO
Foss Infratec				
Instrument Type	Infratec 1241 Grain Analyzer	Infratec 1229 Grain Analyzer	Infratec 1229 Grain Analyzer	Infratec 1225 Grain Analyzer
Serial Number	0350	553792	3108	0065
Technology	Scanning monochromator working in transmittance mode			
Spectral range	850 – 1048 nm	850 – 1048 nm	850 – 1048 nm	850 – 1048 nm
Sampling Interval	2 nm	2 nm	2 nm	2 nm
Instrument Role	Master	Slave	Slave	Slave
Number of Subsamples	10	10	10	10
Used for Calibration	YES	NO	NO	NO

Bruins Instruments

The Bruins Instruments OmegAnalyzer G and AgriCheck, are near infrared transmittance (NIT) grain analyzers operating in wavelength range from 730 to 1100 nm (Table 3). The sampling rate of the instruments is 0.5 nm (741 data point–spectra).

The OmegAnalyzer G and AgriCheck have a built in fully functional Pentium class computer with a keyboard, display, and sample drawer at the front and a power switch and connectors at the rear. Analyzers are calibrated with the GRAMS/32 software (Thermo Galactic, www.thermo.com). The Omega prediction software (Bruins Instruments , www.bruins.de) is a menu oriented graphical user interface which provides all instrument functions. The instrument uses a Windows 98 operating system with a built in RJ45 connector that enables a connection to a LAN or the Internet. Instruments also have two USB and serial ports.

Foss Infratec

The Infratec analyzers are near infrared transmittance (NIT) instruments widely used in the grain industry. Foss 1229 operates from 850 to 1048 wavelength range. Foss 1241 also has an extension module that will extend the operating range and allow a measurement range from 570 to 1100 nm. The sampling rates of the instruments are 2 nm (100 or 265 data points).

The FOSS Infratec 1229 and 1241 use a proprietary operating system with text based user interface. Data exchange is limited to the use of a slow modem connection, serial port, and floppy drives. Foss Infratec 1241 and 1229 also come with a built in PC compatible computer, but because of the proprietary operating system, only serial port communication can be utilized for networking of the instrument. This can be done by using a modem or a RS232 serial port.

SELECTION OF STANDARDIZATION SAMPLES

The first step in developing a standardization transform model is to select samples that will be measured on the machines to be standardized. One method for choosing samples is based upon the multivariate leverage of the samples, which is a measure of their uniqueness in the calibration set. This is a simple procedure which starts by selecting the sample with the greatest deviation from the multivariate mean of the calibration samples. All other samples are then orthogonalized with respect to the first sample and the procedure is repeated. Orthogonalization will help select only samples that varies in different directions than already selected samples.

Given a calibration set from Instrument 1, $\mathbf{R}_1$ (*n* samples by *m* wavelengths), that has been mean centered, calculate the leverage matrix **H** as : $\mathbf{H}=\mathbf{R}_1\mathbf{R}_1^T$ (note that **H** is *n* by *n).* The diagonal elements of **H** are the leverages of the samples (h_{ii} is the leverage of i^{th} sample).

The twenty samples with the highest leverage r_{max} were selected. Orthogonalization of each remaining spectra r_i in the data set is performed by using the following[10]:

Equation 1

$$r_{io}=r_i-r_{max}((r_{max}r_i^T)/(r_{max}r_{max}^T))$$

where: r_i = vector from spectra set

SOFTWARE

To retrieve data from Bruins and Infratec instruments, custom made client software developed in Microsoft Visual Studio™(Microsoft™, msdn.microsoft.com) was used.

Multivariate analysis was performed in a Matlab® version 6.5 (The Mathworks, www.mathworks.com) computational environment. The PLS_Toolbox 3.0 (Eigenvector Research Inc., www.eigenvector.com), Neural Network 4.0.1 Toolbox (The Mathworks, www.mathworks.com), and Statistics Toolbox 4.1 (The Mathworks, www.mathworks.com) were used.

STANDARDIZATION METHODS

PDS Piecewise Direct Standardization

Perhaps the most successful technique currently used for optical standardization is the Piecewise Direct Standardization (PDS) method[11,12,13]. PDS works by forming local linear models that relate the response of the instrument to be standardized over a range of frequencies to the response of the standard instrument at a single frequency.[10] It is assumed that the calibration model is formulated as the following:

Equation 2

$$y = R_1\beta + 1b_1$$

where: **y** is the concentration vector of the property of interest

R_1 is the response matrix

β is the regression vector

1 is a vector of ones with a length equal to the number of samples

After mean centering:

Equation 3

$$y = R_1\beta$$

The relationship between instruments is then modeled as:

Equation 4

$$S_1 = S_2 F_b + 1 b_s^T$$

where:

S_1	is the response matrix of Instrument 1 to the transfer samples
S_2	is the response matrix of Instrument 2
F_b	is the transformation matrix
b_s^T	is the transposed background correction matrix which accommodates the additive background differences between instruments
1	is a vectors of ones of length equal to the number of transfer samples

The transfer function matrix, F_b, is then mean centered and calculated to satisfy Equation 5.

Equation 5

$$S_1 = S_2 F_b$$

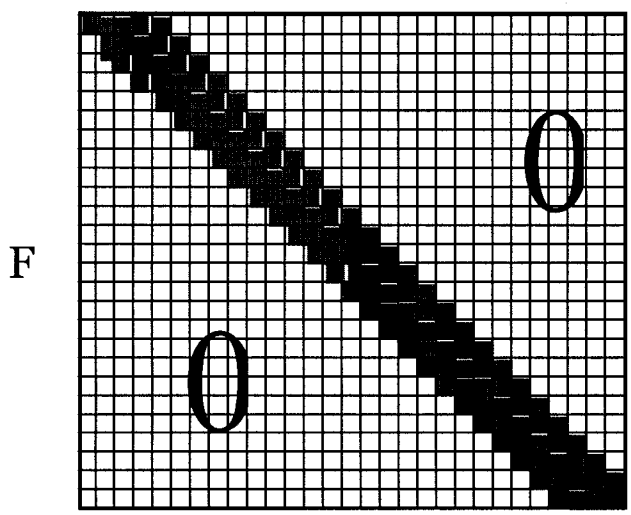

Figure 1. Structure of the PDS Transformation Matrix F_b. Shaded boxes represent wavelengths (x) used in the standardization of wavelengths x_i

Direct Standardization

Direct standardization can be implemented using a univariate or multivariate technique.[14] In this project, a multivariate approach was used. The multivariate approach uses a PDS algorithm, but in order to find the transfer function F_b, it utilizes the whole spectrum instead of a window of a certain size. If baseline photometric difference is expected, bias can also be incorporated.[12]

Multiplicative Scatter Correction

MSC is a spectra processing step that attempts to account for differences in measurement path lengths.[15] The MSC regress measured spectra against reference spectra and then correct the measured spectra using the slope adjustment.

The procedure for MSC is as follows:

First, the s is defined as a column vector corresponding to a spectra to be standardized and r is defined as a vector corresponding to reference spectra (mean spectra of the calibration set). The unknown multiplicative factor b is determined using Equation 6:

Equation 6

$$b = ((r - \bar{r})^T (r - \bar{r}))^{-1} (r - \bar{r})^T (s - \bar{s})$$

where:

r is the reference spectra

s is the spectra to be standardized

$\bar{r}$ is the reference spectra mean

$\bar{s}$ is the spectra to be standardized mean

The corrected spectra are then:

Equation 7

$$s_{corrected} = \frac{(s - \bar{s})}{b} + \bar{r}$$

Finite Impulse Response (FIR) filters

FIR standardization can be described as a moving window that uses MSC and is analogous to using a finite impulse response modeling methodology.[16] The FIR filters use a windowed MSC to correct the spectra to reference spectra with only the center channel of each window being corrected.

This technique has the advantage of using MSC on only one spectrum (i.e., the mean spectrum) for standardization of the second instrument. Also, if the scattering effect varies as a function of the wavelengths, the windowing may avoid some of the limitations imposed by using a single multiplicative factor in MSC.

CALIBRATION ALGORITHMS

Two processing algorithms were implemented in this test:

- a linear regression model (Partial Least-Squares Regression: PLS),
- a non-linear model (Artificial Neural Networks: ANN).

There is no restriction on the number of models or processing algorithms that could be used.

Model Comparison

The Standard Error of Prediction (SEP) was calculated from a verification set of 20 samples with wet chemistry references provided by Woodson–Tenent Laboratories, Inc. (Des Moines, IA). These samples had replicated chemistry values and were laboratory transfer standards. The SEP is the Standard Deviation (SD) of the difference between predicted and reference values.

The SEP was calculated by the following equation:

Equation 8

$$SEP = \sqrt{\frac{\Sigma(y-x)^2 - \frac{(\Sigma(y-x))^2}{N}}{N-1}}$$

where:

y is the result from the chemical analysis

x is the result predicted from NIR measurements

N is the number of samples in the validation set

Ratio of Standard Error of Prediction (SEP) to Standard Deviation (SD), called RPD, was also calculated. The RPD is a statistic that evaluates SEP in terms of the SD of the reference data. If the SEP is similar to the SD of the reference data (RPD $\approx$ 1), the instrument is not predicting the reference data at all.[17] RPD will be the target statistic (Table 4).

Table 4. Guidelines for interpretation of RPD[16]

RPD value	Classification	Application
0.0 – 2.3	Very poor	Not recommended
2.4 – 3.0	Poor	Very rough screening
3.1 – 4.9	Fair	Screening
5.0 – 6.4	Good	Quality control
6.5 – 8.0	Very good	Process control
8.1 +	Excellent	Any application

The RPD was calculated by the equation:

Equation 9

$$RPD = \frac{SD_x}{SEP}$$

where SD_x is Standard deviation of x (optical data) calculated by:

Equation 10

$$SD_x = \sqrt{\frac{\sum x - \frac{(\sum x)^2}{N}}{N-1}}$$

where:

x is the result predicted from NIR measurements

N is the number of samples in the validation set

Partial Least Squares (PLS)

PLS reduces the number of variables by calculating linear combinations of the original variables (factors) and using a small enough number of these factors to allow for a matrix inversion.[3] Both **X** (spectra) and **y** (reference data) are actively used in the data

analysis. This helps to avoid the potential effects of x variables having large variances which are irrelevant to the calibration model.[18,19]

A disadvantage of PLS is that rank determination (determining how many latent variables to use in the model) is not straightforward. The number of latent variables is the parameter that needs to be optimized in order to avoid under or over fitting when using the PLS regression model. In this project, the number of latent variables varied from 6 to 15. Optimization was done by minimizing SEP of the calibration for different number of the latent variables.

Artificial Neural Networks (ANN)

The ANN calibration model is able to fit non-linear relationships between multivariate x and y values. The most common problem that occurs during neural network training is over fitting of the model. Over fitting refers to the situation when the error on the calibration set is optimized to a small value, but will become an unacceptably large when new data are predicted by the model. When overfitting, the calibration is working as look-up table, the training samples are memorized, and the generalization rules are not learned.

Regularization and early stopping are two methods that are commonly used to prevent overfitting. In the early stopping technique, the available data are divided into two subsets. The second set, called the validation set is not included in the calibration. Instead, it serves to monitor calibration training and determine when adjustment parameters for ANN should be stopped in order to avoid overfitting.

Another method for improving generalization is regularization. Regularization implemented in the ANN Toolbox is the Bayesian framework of David MacKay.[20] The

Bayesian framework does not require a validation data set and allows determination of the optimal regularization parameters in an automated fashion. In Bayesian framework the neural network learning is interpreted as an interference of the most probable parameters for the model. The implementation of the Bayesian regularization works best when the network and targets are approximately in the range [-1,1].[21] Implementation of the Bayesian regularization in the ANN toolbox usually provides better generalization performance than early stopping when training for networks function approximation.[20]

The Bayesian regularization was used in the project. Being able to automate ANN training was an important decision factor because in this project several hundred ANNs needed to be evaluated. The ANN toolbox offers many transfer functions that can be used in the neural network design. For hidden layer sigmoid *tansig* transfer function with output in the range [-1,1] was selected.

The two-layer ANN layout of neurons is the ANN design most used for chemical applications. More than 80% of all ANN applications in chemistry use this type of network architecture, although with different numbers of nodes at each level.[22]

The ANN presented in this paper are two layer networks (Figure 2). The weights used in ANN are adjusted by using feed-forward training to model the relationship between descriptors and responses in a supervised learning mode. This type of neural networks is commonly used for multivariate calibration.[23]

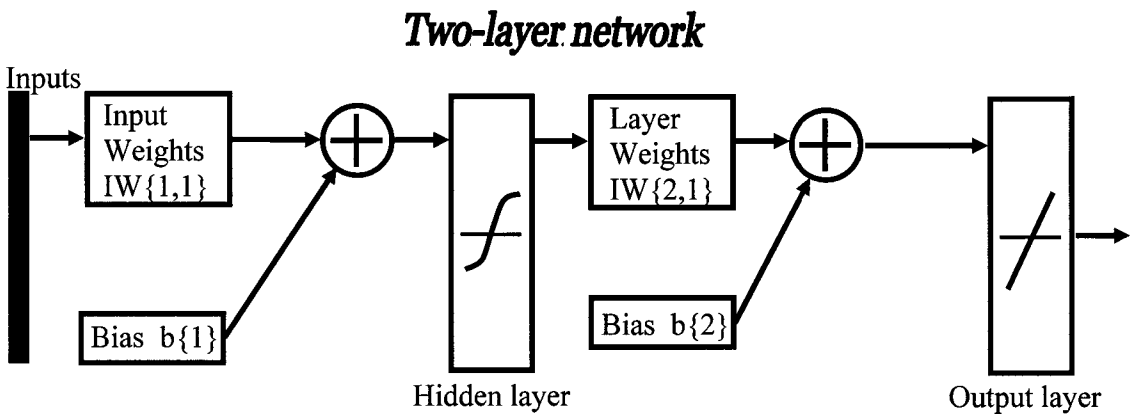

Figure 2. Network architecture used in the project

RESULTS AND DISCUSSION

Results of the between brand instruments standardization transfer for protein and oil predictions are listed in Tables 5, 6, and 7. Calibration for all instruments listed in Tables 5, 6, and 7 were developed on Foss Infratec 1241 Grain Analyzer and then transferred to one Bruins Instruments Agricheck and two Bruins Instruments OmegAnalyser G analyzers.

Table 5 lists the results achieved when the ANN calibration regression model was used. Tables 6 and 7 list the results from the same reference and spectral data. The only difference is that instead of the ANN, the PLS calibration regression model was applied.

Results listed on the first line in Table 5 for FOSS 0350 lists the validation set statistics. None of the other instruments listed in the table were included in the calibration development and therefore, can be used as an independent set to evaluate the performance of the standardization transfer.

Table 5. Direct standardization transfer of ANN calibration model between Foss Infratec and Bruins Instruments for corn protein and oil prediction.

PROTEIN	Protein			Oil		
ANN model with Direct Standard.	R2	RPD	SEP	R2	RPD	SEP
FOSS (master) 0350	0.986	8.7	0.20	0.978	6.3	0.22
BRUINS 6110	0.979	6.9	0.27	0.974	5.6	0.22
BRUINS 6118	0.978	6.5	0.29	0.979	6.3	0.19
BRUINS 6175	0.977	6.6	0.29	0.981	6.4	0.19
BRUINS 31002	0.976	6.3	0.30	0.979	6.3	0.20

Table 6. Direct standardization transfer of PLS calibration model between Foss Infratec and Bruins Instruments for corn protein prediction.

PROTEIN	Protein				
PLS model with Direct Standard.	# PC	RPD	SEP	Slope	Bias
FOSS (master) 0350	12	6.2	0.318	0.978	0.18
Bruins 6110	12	6.3	0.296	0.972	0.23
Bruins 6118	12	6.0	0.311	0.971	0.24
Bruins 6175	12	6.0	0.311	0.971	0.24
Bruins 31002	12	6.1	0.308	0.972	0.23

Table 7. Direct standardization transfer of PLS calibration model between Foss Infratec and Bruins Instruments for corn oil prediction.

PLS model for oil	# PC	RPD	SEP	Slope	Bias
FOSS (master) 0350	11	6.1	0.19	1.010	-0.05
Bruins 6110	11	6.1	0.20	1.016	-0.07
Bruins 6118	11	6.6	0.19	1.018	-0.08
Bruins 6175	11	7.0	0.17	1.020	-0.09
Bruins 31002	11	7.3	0.17	1.021	-0.09

The calibration model developed for Foss Infratec cannot be directly applied to spectral x values of Bruins Instruments analyzers because the instruments operate in different spectral range and with a different sampling rate. A different spectral range was corrected by limiting the spectral range of the Bruins Instruments analyzers to include only 850 – 1048 nm range. The different sampling rate was addressed with a non-square transformation matrix when applying the Direct Standardization transfer. Figures 3 and 4 demonstrate that successful protein and oil calibration transfer between Foss Infratec 1241 Grain Analyzer spectrometers and Bruins Instruments analyzers is achievable.

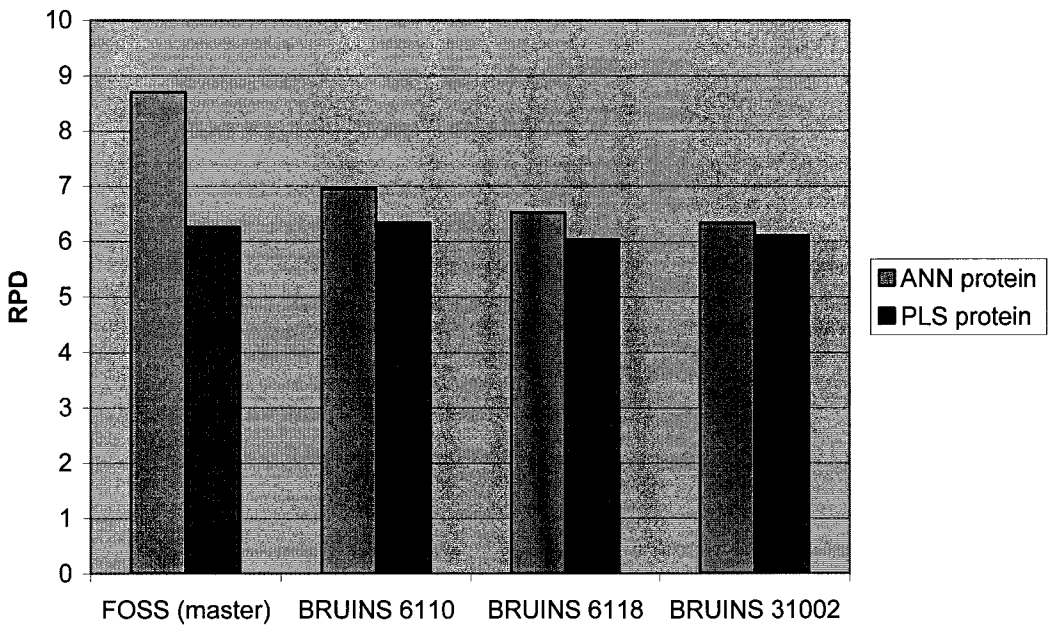

Figure 3. Optical direct standardization of three Bruins Instruments analyzer using calibration developed on Foss 1241 Infratec Grain Analyzer for corn protein prediction

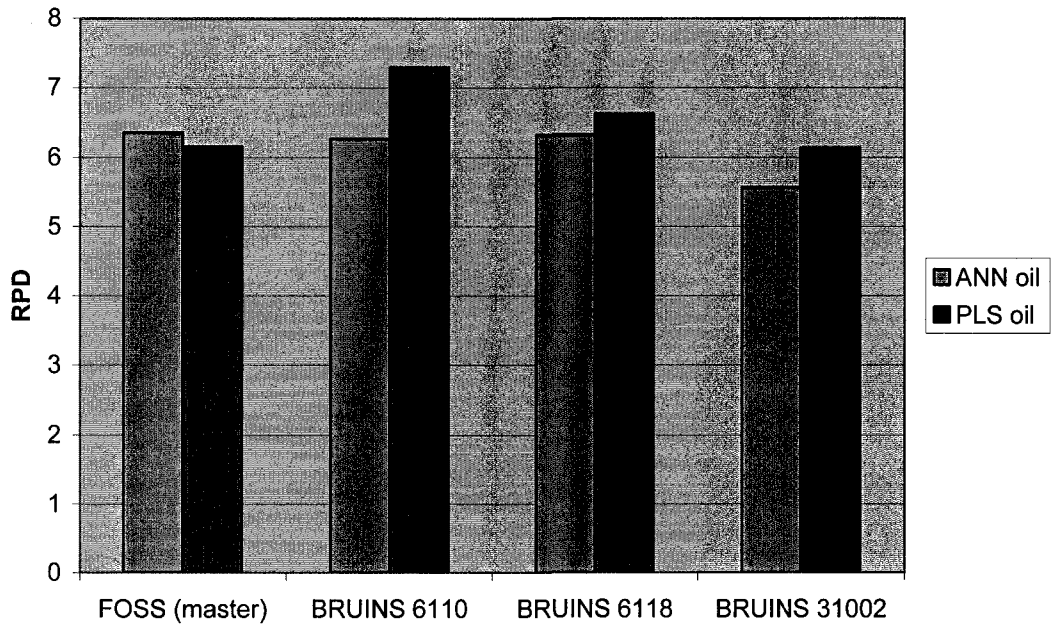

Figure 4. Optical direct standardization of three Bruins Instruments analyzer using calibration developed on Foss 1241 Infratec Grain Analyzer for corn oil prediction

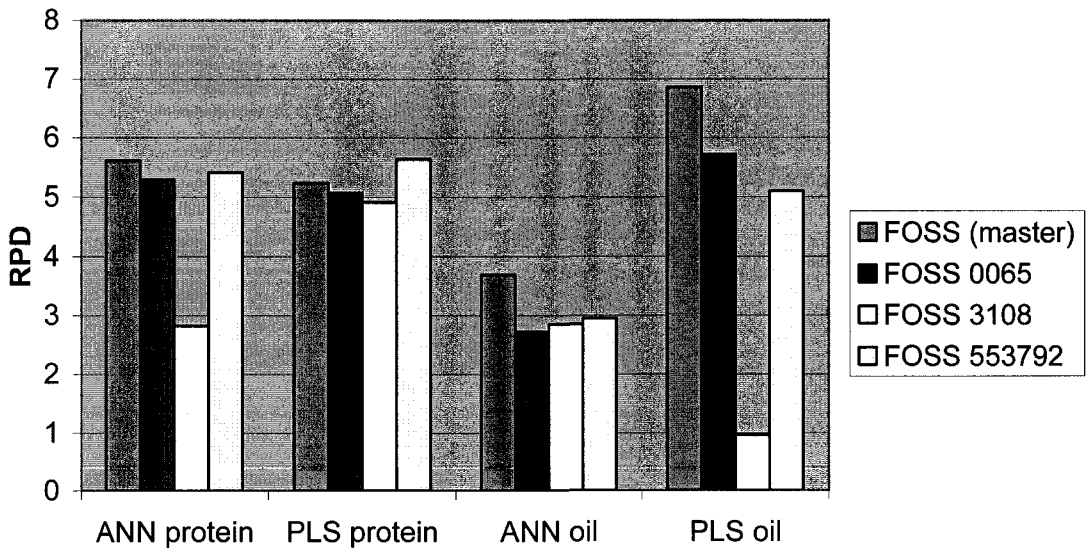

Figure 5. Calibration transfer among Foss Infratecs using Finite Impulse Response (FIR) standardization, corn

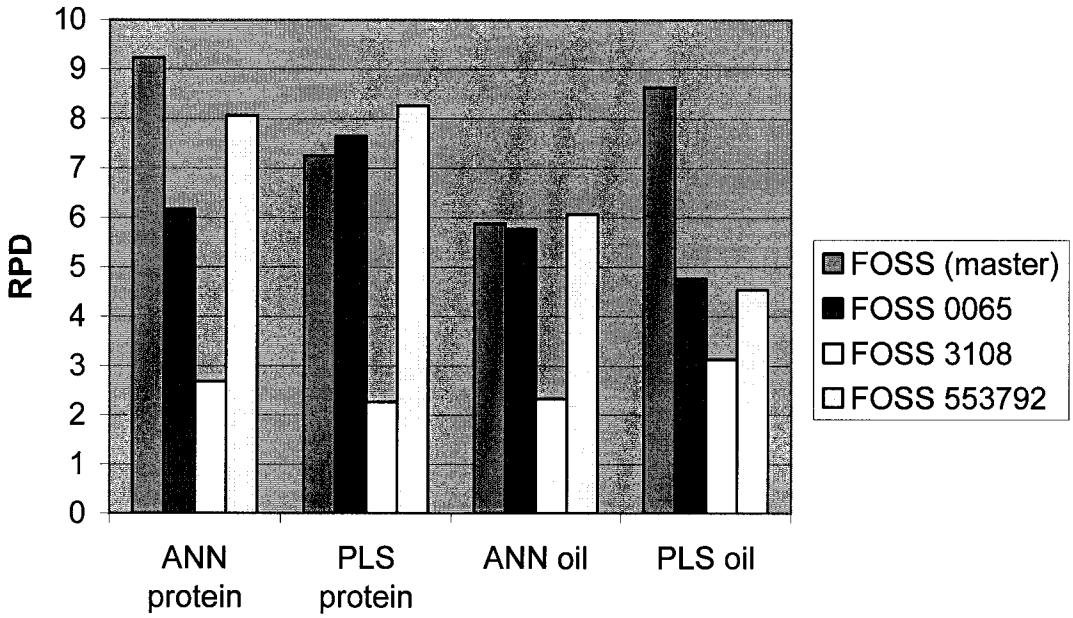

Figure 6. Calibration transfer among Foss Infratecs with Multiplicative Scatter Corrections (MSC) pretreatment, corn

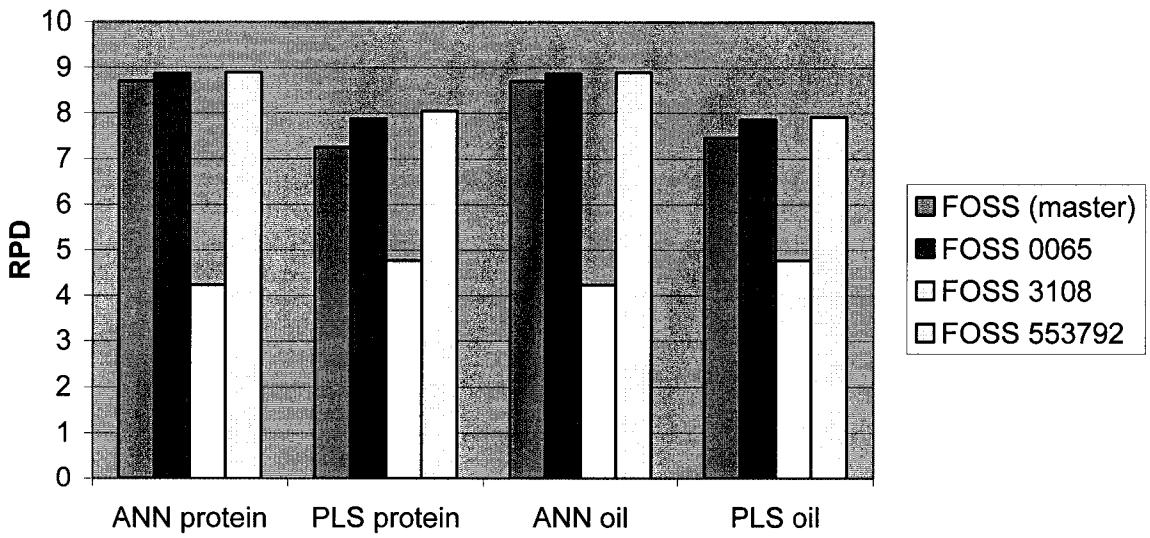

Figure 7. Calibration transfer of Foss Instruments without standardization

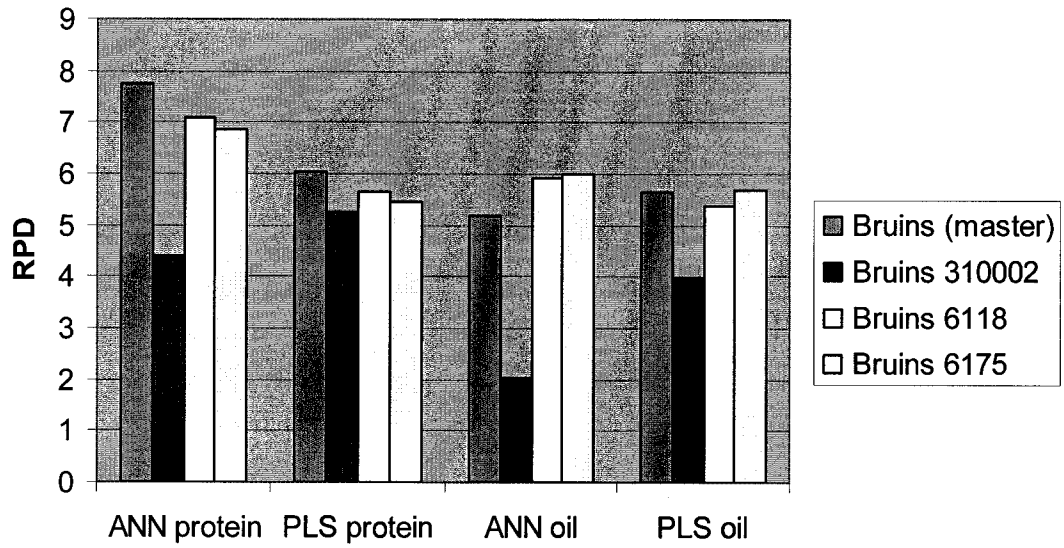

Figure 8. Calibration transfer of Bruins Instruments analyzers without using standardization transfer

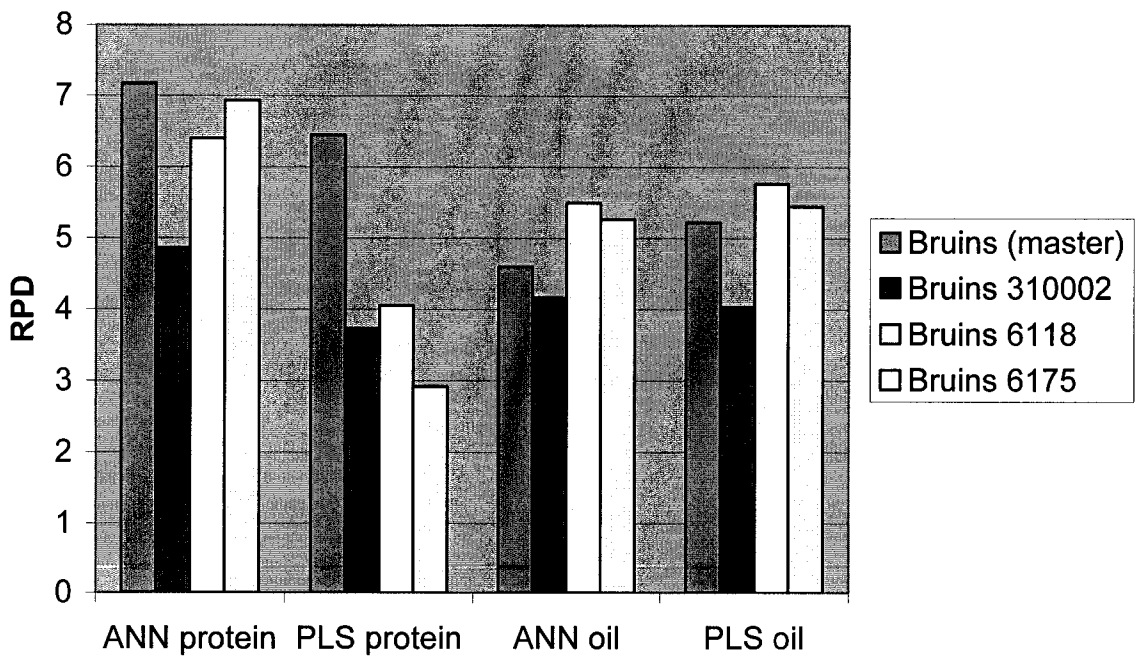

Figure 9. Calibration transfer of Bruins Instruments analyzers by using Multiplicative Scatter Corrections (MSC) pretreatment

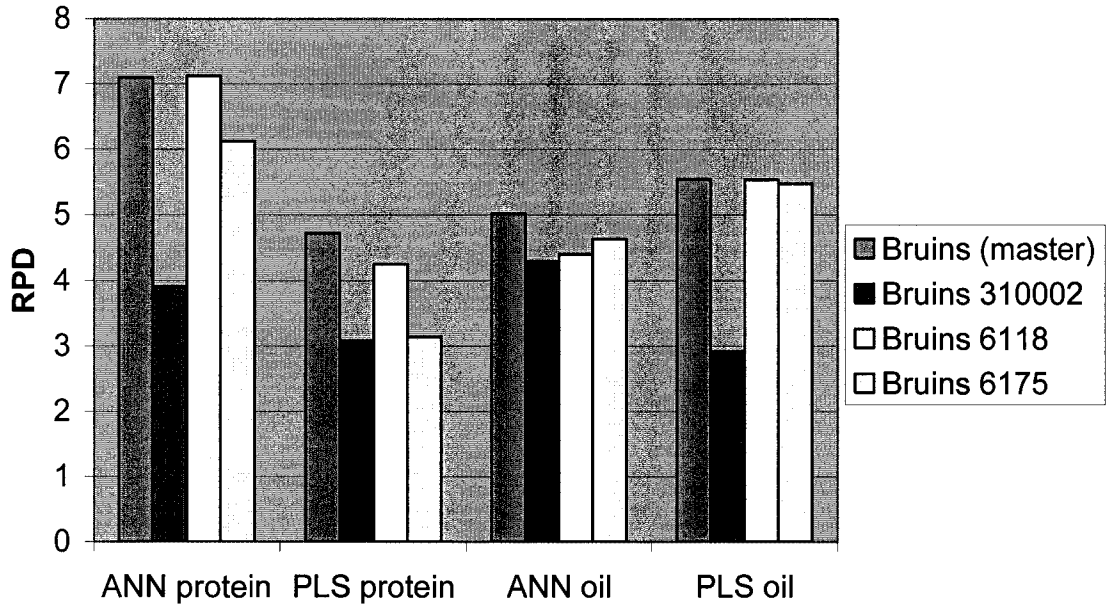

Figure 10. Calibration transfer among Bruins Instruments analyzers with Finite impulse Response (FIR) standardization transfer

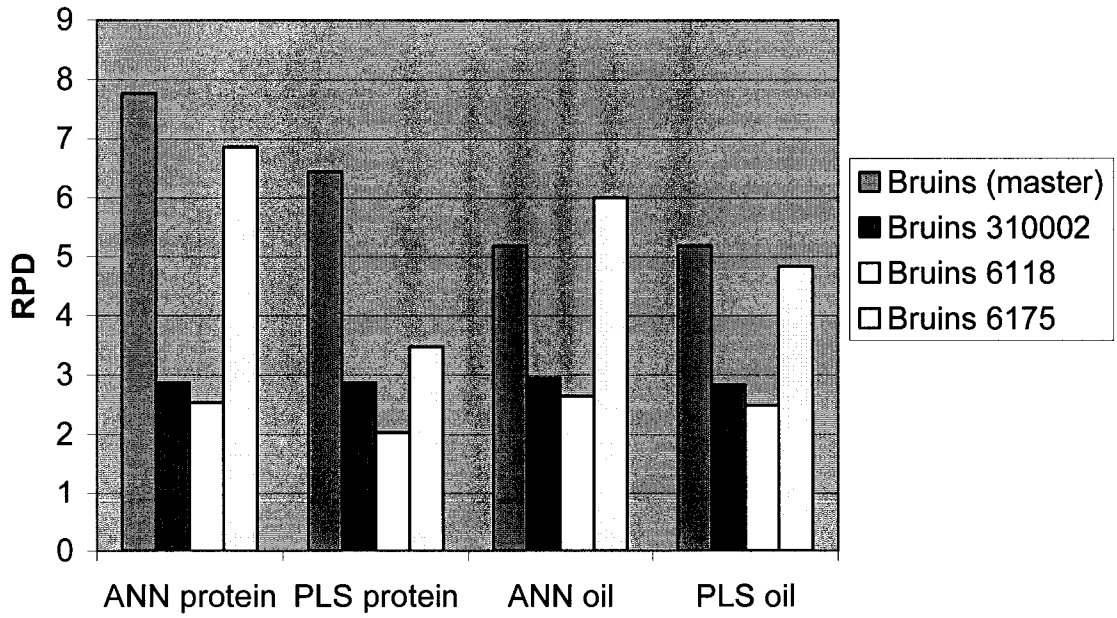

Figure 11. Calibration transfer among Bruins Instruments analyzers with Piecewise Direct Standardization (PDS) transfer

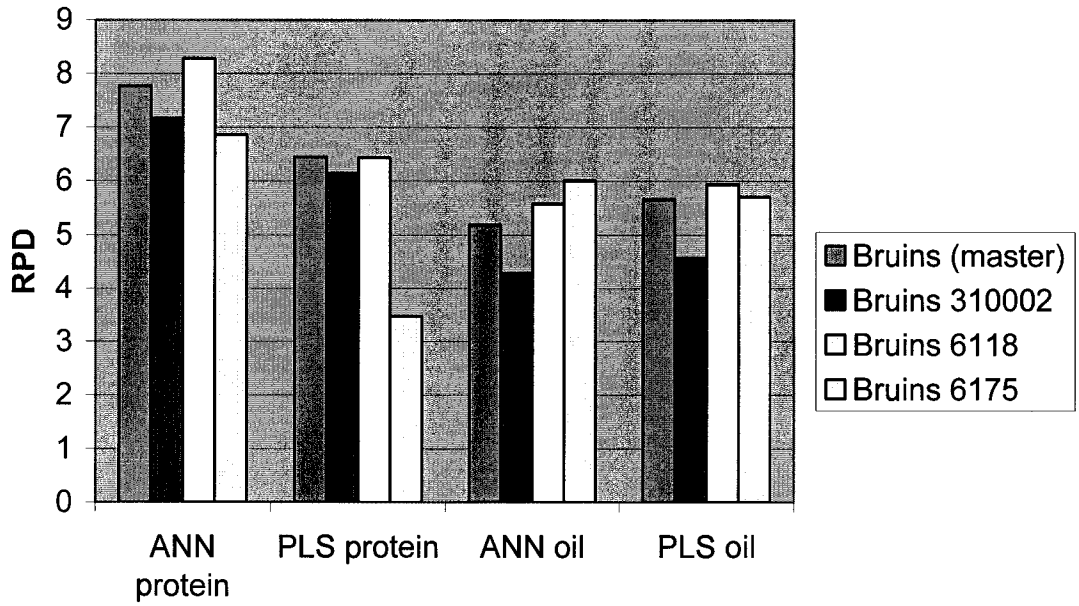

Figure 12. Calibration transfer of Bruins Instruments analyzers with Direct Standardization transfer

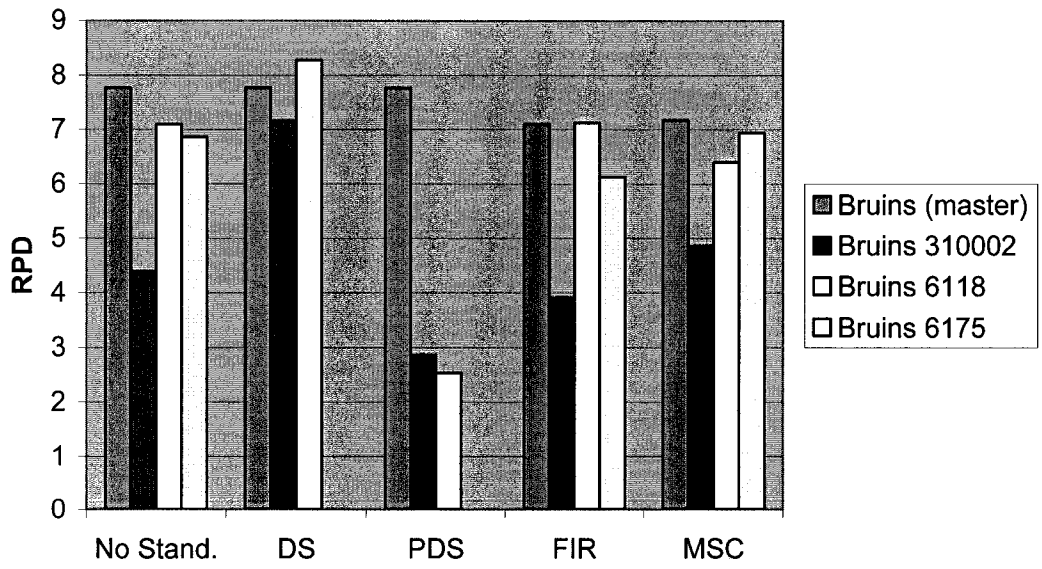

Figure 13. Comparison of different standardizations developed for Bruins Omega, ANN corn protein calibration

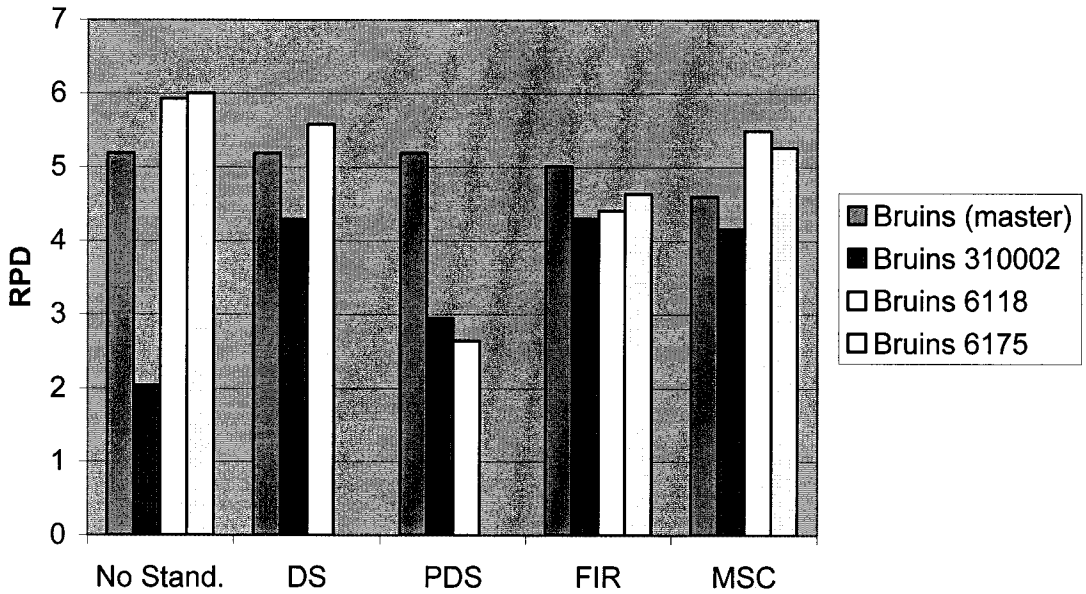

Figure 14. Comparison of different standardizations developed for ANN corn oil calibration

Figures 5 and 6 compares the results of two standardization approaches calculated on a group of one master and three slaves Foss Infratec Grain Analyzer instruments. Figure 7 compares the results when no standardization was applied. Calibration was developed on Foss Infratec 1241 Grain Analyzer displayed on the plots as Foss (master). Slave instruments were one Foss Infratec 1229 Grain Analyzer and two Foss Infratec 1229 Grain Analyzer instruments.

RPD was calculated by using an independent prediction set of 20 samples. Only MSC and FIR standardization were used to standardize Foss Infratec spectra. Figure 8 compares the results when no standardization was applied using Bruins Instruments analyzers. Bruins Instruments analyzers standardization results are compared in Figures 9 – 14. Calibrations for Bruins analyzers were developed on one OmegAnalyzer G with serial number 6110, displayed on the plot legend as Bruins (master). RPD values were calculated by using an

independent prediction set of 20 samples. PDS, DS, MSC and FIR standardizations were used to standardize Bruins Instruments analyzers. The best results were achieved when the DS optical standardization method was used. Both Bruins and Foss units used the same calibration set of 480 corn samples. MSC and FIR standardization methods (Figures 6,7,9 and 10) did not improve prediction.

CONCLUSIONS

A number of published near-infrared studies are dedicated to the problem of standardization transfer.[1,24,25,26] Several standardization transfer approaches were explored in this study.

OmegAnalyzer G instruments from Bruins Instrument analyzers were successful in implementing a calibration developed on the master OmegAnalyzer G. Both slave OmegAnalyzer G instruments (Serial number 6118 and 6175) can be used in the instrument network to predict corn protein and oil content without standardization (Figure 7 and 8). For the third Bruins Instruments AgriCheck with serial number 310002, performance improved after the Direct Standardization transfer was applied. For this instrument, RPD value for protein prediction using ANN calibration improved from 4.4 to 6.3 (Figure 13). Similarly, for oil prediction, RPD value changed from 2.0 to 4.3 (Figure 14).

For the Foss Infratec grain analyzers, only MSC and FIR standardization were used. The Foss Infratec 1229 with serial number 3108 proved to be most difficult to standardize. Both methods failed to improve RPD values of the Foss Infratec 1229 with serial number 3108. After applying MSC and FIR standardization RPD value dropped below 3. Better

results were achieved when MSC and FIR standardization was applied to data from Bruins Instruments analyzers. FIR standardization and MSC preprocessing are very desirable methods for standardization because they do not require standardization samples. This is important especially for large network of instruments.

In this research PLS and ANN models were used. Relatively small number of calibration samples favored the linear PLS model, but measured results show that the non-linear ANN can perform well even if the number of samples is less than 500 (Figure 6–9, 12). Even if less than 500 samples with reference data is used for calibration, development of NIRS calibrations for agricultural commodities is very costly. Therefore, the possibility of using same calibration between different NIR analyzer manufacturers was proven feasible.

Direct standardization was a sufficient method to successfully predict protein and oil content of corn using Bruins Instruments OmegAnalyzer G and AgriCheck analyzers with ANN calibrations developed on Foss Infratec 1241 Grain Analyzer. Transfer was successful, recorded RPD values were above 6 for protein and above 5 for oil predictions (Figures 3 and 4). Currently, only the Foss Infratec Whole Grain Analyzer manufactured by Foss North America, Inc., has been approved by the USDA's Grain Inspection, Packers and Stockyards Administration (GIPSA) for official determination of protein content in wheat; protein and oil content in soybeans; and protein, oil, and starch content in corn. Successful transfer of the calibration models from Foss Infratec Whole Analyzers to other near infrared transmittance analyzers will allow GIPSA and official inspection agencies the opportunity to purchase new near infrared spectroscopy equipment that is approved for official inspection purposes.

REFERENCES

[1] P. Dardenne, "Calibration transfer in near infrared spectroscopy", Near Infrared Spectroscopy: Proceeding of the 11th International Conference, Edited by A.M.C Davies and A. Garrido-Varo

[2] Kurt Varmuza, "Multivariate Data Analysis in Chemistry", "Handbook of Chemoinformatics – From Data to Knowledge" , Edited by Johann Gasteiger, Copyright © WILEY-VCH Verlach GmbH, 2003, p.1117

[3] Kenneth R. Beebe, Randy J. Pell, May Beth Seasholtz, "Chemometrics A Practical Guide", A Wiley-Interscience Publication,1998

[4] H.W. Siesler, S.Kawata, H.M.Heise "Chemometrics in Near-Infrared Spectroscopy", "Near-Infrared Spectroscopy. Principles, Instruments, Applications", Copyright © WILEY-VCH Verlach GmbH, 2002, p.154

[5] J. Siska, "Standardization of near-infrared transmittance instruments", © 1997, Iowa State University, p.36

[6] C.V. Greensill, P.J. Wolfs, C.H. Spiegelman, and K.B. Walsh, "Calibration Transfer between PDA-Based NIR Spectrometers in the NIR Assessment of Melon Soluble", Applied Spectroscopy, © 2001 Society for Applied Spectroscopy, Volume 55, Number 5, 2001

[7] Hartwig Schultz and Sven Pfeffer, "Calibration transfer applied for selected medicinal plants the Shenk-Westerhaus algorithm", Near Infrared Spectroscopy: Proceeding of the 11th International Conference, Edited by A.M.C Davies and A. Garrido-Varo

[8] Robert Dzupin, Sylvie A. Rousell, Glen R. Rippke, and CharlesR. Hurburgh, Jr., NIR_GrainNET: Internet-enabled software for real-time prediction using NIR whole-grain analyzers, PITTCON, 2002

[9] Ching-Hui Tseng, Nan Wang, Kangming Ma, "A successful story of an Internet-enabled NIR system", Near Infrared Spectroscopy: Proceeding of the 11th International Conference, Edited by A.M.C Davies and A. Garrido-Varo, p.29

[10] Barry M. Wise, Neal B. Gallagher, Rasmus Bro, and Jeremy M. Shaver, "PLS_Toolbox 3.0 for use with MATLAB™", ©2002 Eigenvector Research, p.131-3

[11] Wang et al., "Multivariate Instrument Standardization," Anal. Chem., Vol. 63, p. 2750-2756 (1991)

[12] Wang et al., "Improvement of Multivariate Calibration through Instrument Standardization," Anal. Chem., Vol. 64, pp. 562-584 (1992)

[13] Wang et al., "Additive Background Correction in Multivariate Instrument Standardization," Anal. Chem., Vol. 67, pp. 2379-2385 (1995)

[14] Tormod Naes, Thomas Isakson, Tom Fearn, Tony Davies, "A user-friendly guide to Multivariate Calibration and Classification", © 2002 NIR Publications, p.209

[15] H. Martens, and T. Naes, Multivariate Calibration, John Wiley & Sons, New York, NY, 1989

[16] T. B. Blank, S.T. Sum, S.D. Brown, and S.L. Monfre, "Transfer of Near-Infrared Multivariate Calibrations without Standards", Anal. Chem., 68(17), 2987-2995, 1996

[17] P.C. Williams, "Implementation of Near-Infrared Technology", Near-Infrared Technology in the Agricultural and Food Industries", Edited by Phil Williams and Karl Norris, ©2001 by the American Association of Cereal Chemists, Inc.,p. 164

[18] M.J. Adams, "Chemometrics in Analytical Spectroscopy 2nd Edition, RSC Analytical Spectroscopy Monographs, Published by Royal Society of Chemistry 2004, p. 203

[19] Lennart Eriksson, Henrik Antti, Elaine Holmes, and Erik Johansson , "Partial Least Squares (PLS) in Cheminformatic", "Handbook of Chemoinformatics – From Data to Knowledge" , Edited by Johann Gasteiger, Copyright © WILEY-VCH Verlach GmbH, 2003, p.1134

[20] Howard Demuth, Mark Beale, "Neural Network Toolbox – For use with MATLAB®", ©2000 by The Math Works, Inc.

[21] David John MacKay, "Bayesian Methods for Adaptive Models", © 1992 California Institute of Technology.

[22] Jore Zupan, "Neural Networks", "Handbook of Chemoinformatics – From Data to Knowledge" , Edited by Johann Gasteiger, Copyright © WILEY-VCH Verlach GmbH, 2003, p.1173

[23] Frederic Despagne, Beata Walczak, and Desire-Luc Massart, "Transfer of Calibrations of Near-Infrared Spectra Using Neural Networks", Applied Spectroscopy, © 1998 Society for Applied Spectroscopy, Volume 52, Number 5, 1998, p 733

[24] J. P. Conzen, T. Stadelmann, A. Schmidt, H. Weiler, T. Droz-Georget and A. Zillan, „Calibration transfer – aquestion of your needs", "Near Infrared Spectroscopy: Proceedings of the 9th International Conference", NIR Publications 2000,

[25] Edward Stark and Karen Luchter, "Diversity in NIR instrumentation", Near Infrared Spectroscopy: Proceeding of the 11th International Conference, Edited by A.M.C Davies and A. Garrido-Varo, NIR Publications 2002

[26] P.J. Clancy, "Spectral matching two different brands of NIR whole grain analysers", Near Infrared Spectroscopy: Proceeding of the 11th International Conference, Edited by A.M.C Davies and A. Garrido-Varo, NIR Publications 2002

CHAPTER 4. DEVELOPMENT OF A NEW NIRS CALIBRATION MODEL USING COEFFICIENTS FROM WAVELET DECOMPOSITION IN FEED-FORWARD NEURAL NETWORK ARCHITECTURE

A paper to be submitted in the *Journal of NIR Spectroscopy*

Robert Dzupin and Charles. R. Hurburgh, Jr.

Agricultural and Biosystems Engineering Department, Iowa State University, 1551 Food Sciences Building, Ames, IA 500 14, USA

ABSTRACT

The use of wavelets as a pre-treatment method for an Artificial Neural Networks (ANN) regression model to create robust calibration was explored. By using multiple-level decomposition, wavelet coefficients from several resolution levels were computed. In the next step, de-noising of the analyzed spectra was performed. Based on the assumption that high frequency components of the spectral signal do not contain significant information about the tested property, detail wavelet coefficients were removed. Two ANN regression models were then developed. Using the same spectral and chemistry data, PCA data compression and Wavelet coefficients were used as an input for ANN calibration. Both ANN models were applied to predict crude protein and oil content of the corn samples that had been measured on several Bruins Instruments grain analyzers. No additional standardization transfer was applied. Ratio of Standard Error of Prediction to Standard Deviation, called

RPD, and Standard Error of Prediction (SEP) were calculated. Improvement of the prediction of protein and oil was observed for the instruments that used the Wavelets decomposition as an input to the ANN calibration regression model.

INTRODUCTION

Development of calibration models for near-infrared spectrometers is complicated by the presence of noise in the spectral data. In past decades, a large number of digital filters have been used for the reduction of the noise. Recently, Wavelet Transform (WT) has been identified as an effective method for removing noise from the chemical data. The most successful wavelet functions in chemistry are the Daubechies wavelet series.[1]

The advantage of the WT method is recognized mainly in the multiresolution of data, which is a process of decomposing signals according to frequency. Decomposition of the spectral data by the WT method allows the elimination of background and baseline noise, since the signal being processed usually contains contributions with different localizations and different locations in the wavelengths (time) and frequency domains.[2]

The most common method of dealing with highly collinear spectral data and extracting relevant information in spectral data is the Principal Component Analysis (PCA).[3] The PCA tends to use global features. Wavelet decomposition can be used to locally distinguish between significant features and features associated with noise. Therefore, Wavelet Transform is a promising tools as an alternative to PCA.[4]

Wavelets Transformation can be also used as complement to principal component analysis to remove the low-frequency scales representing low-frequency components of

independent variables such as seasonal fluctuations and other long-term variations, prior to principal component analysis.[5]

OBJECTIVES

- Apply modified wavelet transform as input for the feed-forward back-propagation ANN regression model
- Compare the new model with a traditional ANN that uses Principal Component Analysis (PCA) for spectral data compression to calculate its inputs

MATERIAL AND METHODS

SPECTRA AND CALIBRATION DATA

In total, 510 corn samples were used in this study. All samples had wet chemistry references provided by Woodson–Tenent Laboratories, Inc. (Des Moines, IA). Samples were split into two groups (Table 1). The first group, called the calibration set, was used to develop the two calibration models. The second group was used as the validation set. The validation set consisted of 20 samples. These samples had replicated chemistry values and were laboratory transfer standards (Table 2).

Table 1. Samples set description

	Calibration Set	Validation Set
1996	0	4
1997	0	2
1998	0	2
1999	145	5
2000	137	1
2001	98	5
2002	25	5
2003	75	6
SUM	480	30

Table 2. Data set description

Property	Moisture (%)	Protein (%)	Oil (%)
	Calibration set		
Range	8.1 – 23.0	5.5 – 16.4	2.24 – 13.2
Moisture Base	As is	Direct 15%	Direct 15%
N	480	480	480
	Validation set		
Range	9.1 – 18.8	5.5 – 14.4	3.3 – 8.6
Moisture Base	As is	Direct 15%	Direct 15%
N	30	30	30

INSTRUMENTS

Four Bruins Instruments analyzers were used in this project (Table 3). One OmegAnalyzer G with a serial number of 6110 was used to develop the calibration models. All four instruments were then used to test samples from the validation set. All samples were

measured during a period of two weeks by one operator. The validation set was measured on all tested instruments in one day.

The Bruins Instruments OmegAnalyzer G and AgriCheck are near-infrared spectrometers that analyze the composition of samples using the near infrared absorbance characteristics of the sample spectra. They operate from 730 to 1100 nm range. The sampling rate of the instruments is 0.5 nm (741 data points).

Table 3. Description of instruments used

Bruins Instruments				
Instrument Type	Omeg AnalyzerG	AgriCheck	Omeg AnalyzerG	Omeg AnalyzerG
Instrument Role	Master	Slave	Slave	Slave
Serial Number	6110	31002	6118	6175
Technology	Scanning monochromator working in transmittance mode			
Spectral range	730 – 1100 nm	730 – 1100 nm	730 – 1100 nm	730 – 1100 nm
Spectral range used for measurements	850 – 1048 nm	850 – 1048 nm	850 – 1048 nm	850 – 1048 nm
Sampling Interval	0.5 nm	0.5 nm	0.5 nm	0.5 nm
Instrument Role	Master	Slave	Slave	Slave
Number of Subsamples	16	10	16	16
Used for Calibration	YES	NO	NO	NO

The OmegAnalyzer G and AgriCheck have a built in fully functional Pentium class computer with a keyboard, display, and sample drawer at the front and a power switch and connectors at the rear. Bruins analyzers are calibrated with the GRAMS/32 software (Thermo

Galactic, www.galactic.com). The Omega prediction software is a menu oriented graphical user interface which provides all instrument functions. The instrument uses a Windows 98 operating system with a built in RJ45 connector that enables a connection to a Local Area Network (LAN) or the Internet. Instruments also have two USB and serial ports.

SOFTWARE

Custom made client software developed in Microsoft Visual Studio™ (Microsoft™, msdn.microsoft.com) was used to retrieve spectra from the prediction software. Custom made client software also exported spectra to the MS SQL server database (Microsoft™, www.microsoft.com\sql) and MATLAB (The Mathworks, www.mathworks.com). Quantitative analyses were performed using the MATLAB version 6.5 computational environment with Neural Network Toolbox version 4.0 (Eigenvector Research Inc., www.eigenvector.com) and the Wavelet Toolbox version 2.2 (The Mathworks, www.mathworks.com).

The Wavelet Toolbox is a collection of functions built in the MATLAB® Technical Computing Environment. It provides tools for the analysis and synthesis of signals and images, as well as tools for statistical applications, using wavelets and wavelet packets within the framework of MATLAB®. Matlab scripts developed for wavelet decomposition and client software were integrated to GrainNet software[6].

PROCESSING ALGORITHMS

Two processing algorithms were implemented in this project:

- Selected wavelet coefficients retrieved by using multiple-level decomposition were used as inputs for the feed-forward back-propagation ANN regression model. Only low frequency wavelet coefficients were selected.
- Feed-forward back-propagation ANN regression model with principal components used as inputs

WAVELETS

A wavelet is a waveform of effectively limited duration that has an average value of zero. Wavelet analysis consists of decomposing a signal or an image into a hierarchical set of approximations and details. At each level j, the j-level approximation called Aj is built, or approximation at level j, and a deviation signal called the j-level detail Dj, or detail at level j. The one-dimensional analysis performed in this project is based on one scaling function ϕ and one wavelet ψ.

The general equation for a wavelet transform on the (b,a) half plane is:

$$W(b,a) = \int \Psi^*_{a,b}(\lambda) f(\lambda) d\lambda$$

where:

W(b,a)	is the wavelet transform,
$\Psi^*_{a,b}(\lambda)$	is the basis function,
a	is the scale factor,
b	is the position term, and
f(λ)	is the spectrum. [7]

Multiple level decomposition calculates the wavelet coefficients from spectra data *s* by producing two set of coefficients: approximation coefficients A1, and detail coefficients D1. These vectors are obtained by convolving signal *s* with the low-pass filter to obtain A1, and with a high-pass filter for D1. (Figure 1)

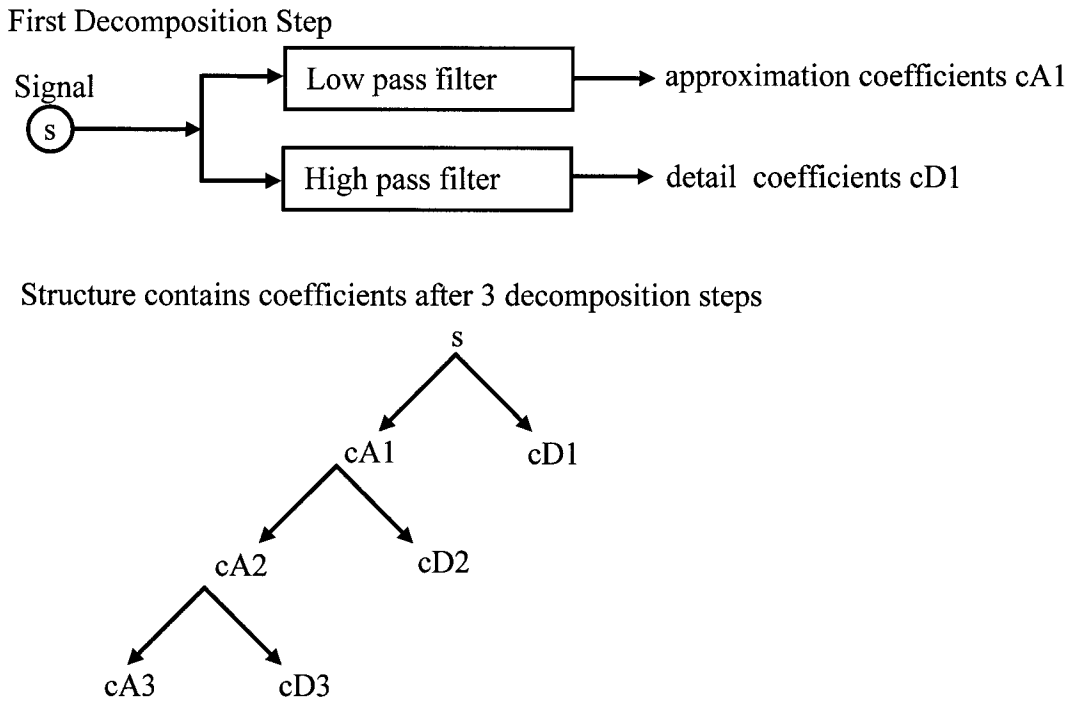

Figure 1. One-dimensional Discreet Wavelet Transform

Discrete Wavelet Transform (DWT) using multiple-level decomposition was performed. The DWT is based on powers of two also called dyadic scales and positions. Twelve types of Daubechies wavelet shapes were used for decomposition. With each wavelet decomposed at second, third, fourth, fifth, and sixth decomposition levels, approximation coefficients were extracted. Extracted coefficients were then used for the development of an ANN calibration on one instrument. Two calibrations with the best Ratio of Standard Error

of Prediction to Standard Deviation (RPD) were then selected. All four instruments were used to predict corn protein and oil content using the 30 validation samples.

ARTIFICIAL NEURAL NETWORK CALIBRATION

Artificial Neural Networks (ANN) calibration model is able to fit non-linear relationships between multivariate x and y-values. The neural networks presented in this paper are two layer networks (Figure 2). The weights used in the ANN are adjusted by using a feed-forward training method to model the relationship between the descriptors and responses in a supervised learning mode. This type of ANN is commonly used for multivariate calibration.[8] To improve generalization, the Bayesian regularization was used in the project.

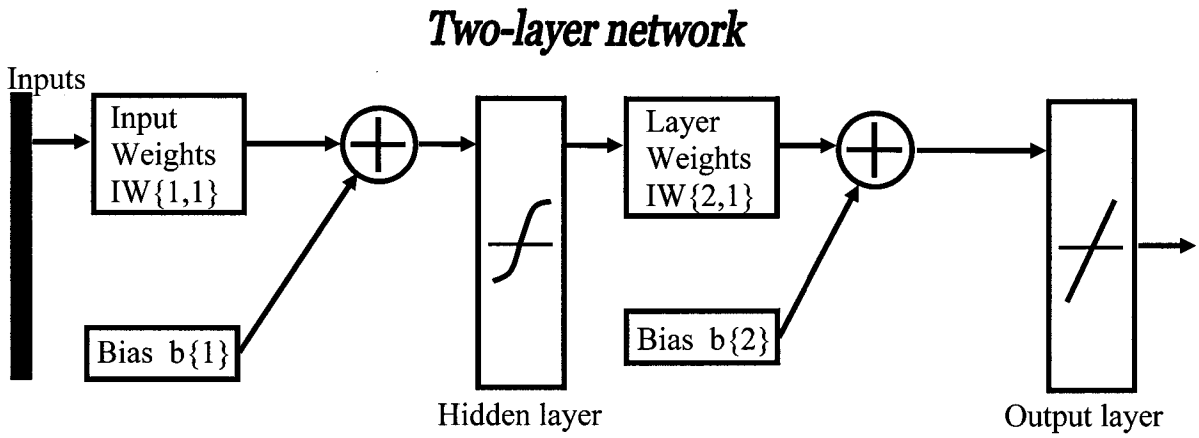

Figure 2. Network architecture used in the project

MODEL COMPARISON

The Standard Error of Prediction (SEP) was calculated from a validation set of 30 samples with wet chemistry references provided by Woodson–Tenent Laboratories, Inc. (Des Moines, IA). SEP was calculated by the following equation[9]:

Equation 1.

$$SEP = \sqrt{\frac{\Sigma(y-x)^2 - \frac{(\Sigma(y-x))^2}{N}}{N-1}}$$

where:

y is the result from the chemical analysis

x is the result predicted from NIR measurements

N is the number of samples in the validation set

Ratio of Standard Error of Prediction to Standard Deviation, called RPD, was also calculated. The RPD is a statistic that evaluates SEP in terms of the SD of the reference data. If the SEP is similar to the SD of the reference data (RPD $\leq$ 1), the instrument is not predicting the reference data.[9]

Equation 2.

$$RPD = \frac{SD_x}{SEP}$$

where SD_x is the Standard deviation of x (optical data) and is calculated as:

Equation 3.

$$SD_x = \sqrt{\frac{\sum x - \frac{(\sum x)^2}{N}}{N-1}}$$

where:

x is the result predicted from the NIR measurements

N is the number of samples in the validation set

RESULTS AND DISCUSSION

Results of the calibration and between instrument calibration transfers are listed in Tables 4 and 5. Bruins Instruments grain analyzer with a serial number of 320002 used a lower number of spectra sub-samples (10 instead of 16) and measured spectra were noticeably noisier than spectra of the other three analyzers (Figure 3).

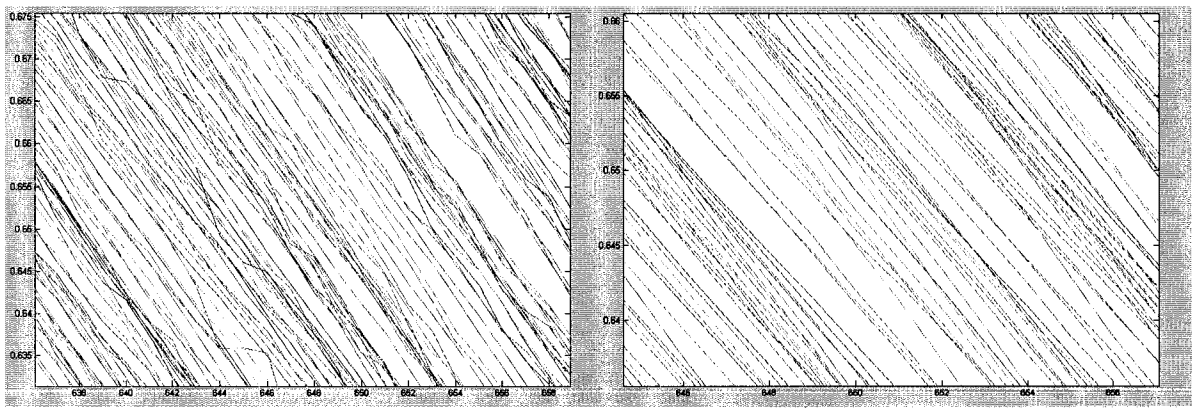

Figure 3. On the left side – a zoomed view on the spectra of the Bruins 310002; on the right side – spectra of the Bruins 6110, using the same set of samples.

Table 4. SEP and RPD for calibration transfer of ANN model in wavelets domain

	RPD			SEP		
CORN PROTEIN	**PCA**	**db6, 5, 23**	**db3, 4, 29**	**PCA**	**db6, 5, 24**	**db3, 4, 30**
Bruins 6110 (master)	7.8	7.2	7.7	0.263	0.283	0.264
Bruins 310002	4.4	4.3	5.3	0.466	0.477	0.387
Bruins 6118	7.1	6.1	6.3	0.292	0.339	0.329
Bruins 6175	6.8	6.3	6.7	0.290	0.316	0.299

Note: db6, 5, 23 means Daubechies wavelets DB6 on 5^{th} level of decomposition using 23 coefficients

Table 5. SEP and RPD for calibration transfer of ANN model in wavelets domain

	RPD			SEP		
CORN OIL	**PCA**	**db6, 5, 23**	**db3, 4, 29**	**PCA**	**db6, 5, 24**	**db3, 4, 30**
Bruins 6110 (Master)	5.2	4.9	4.1	0.230	0.245	0.292
Bruins 310002	2.0	4.3	3.9	0.584	0.276	0.307
Bruins 6118	5.9	5.4	4.4	0.204	0.223	0.276
Bruins 6175	6.0	4.6	5.0	0.194	0.256	0.233

Note: db6, 5, 23 means Daubechies wavelets DB6 on 5^{th} level of decomposition using 23 coefficients

Tables 4 and 5 demonstrate that Principal Component Analysis (PCA) transformation of inputs for ANN can be replaced by using wavelets decomposition. The largest improvement was made for the noisier instrument 310002.

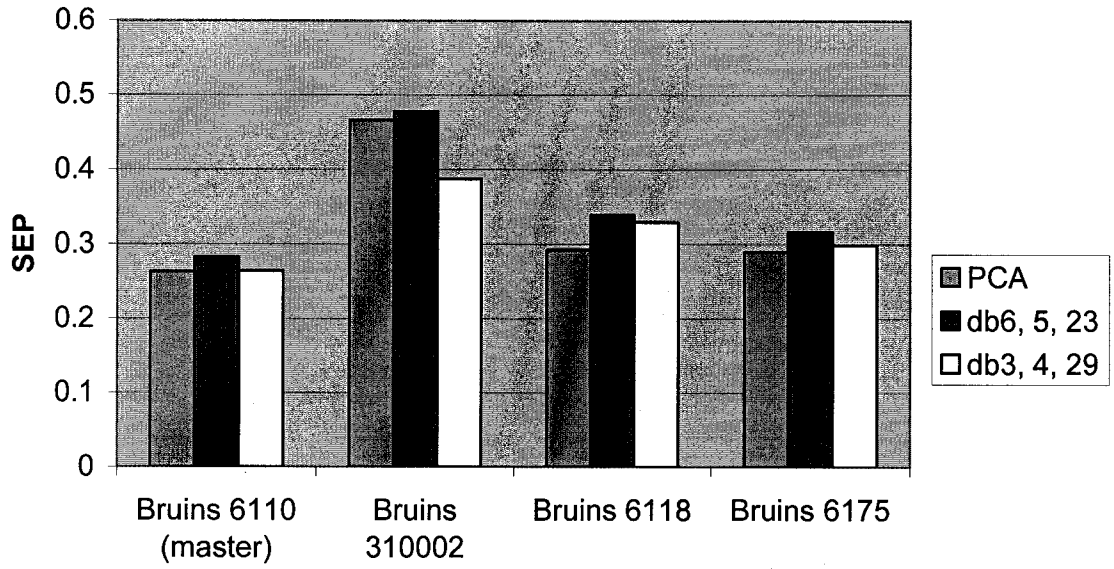

Figure 4. Comparison of PCA and wavelet preprocessing applied to ANN calibration model for protein content in corn

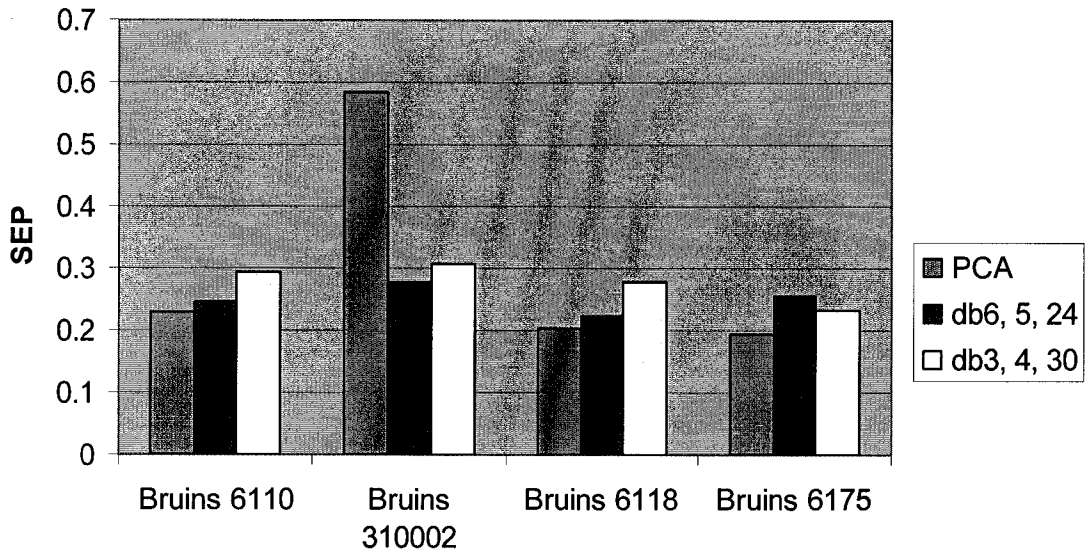

Figure 5. Comparison of PCA and wavelet preprocessing applied to ANN calibration model for oil content in corn

Standard errors of prediction are plotted in Figure 4 and Figure 5. De-noising spectra data by removing detail coefficients in multilevel wavelet decomposition used for the development of the ANN calibration model improved the SEP (robustness) of the developed calibration.

CONCLUSIONS

Multilevel wavelets decomposition in combination with the ANN calibration prediction model improved predictions on the Bruins Instruments AgriCheck analyzer with noisy corn spectra. In this project multilevel decomposition was used not only to compress spectral data but also to selectively remove high frequency components of the spectra after decomposition. Removal of the high frequency part of the noisy spectra improved the standard error of prediction of crude protein content in the corn from 0.47 to 0.39. Even better results were achieved for the prediction of oil content in the corn. Standard error of prediction was improved from 0.58 for PCA – ANN model to 0.31 for Wavelets – ANN model. Wavelets transfer did not improve the standard error of prediction and coefficient of determination when spectra of the instrument were less noisy.

In this project, after wavelet compression, only high frequency components were removed. Therefore, drifts of the baseline, periodic seasonal fluctuations, and long-term drifting, common problems that standardization transfer is designed to address, will not be corrected by using this calibration method. Removal of the additional wavelet coefficients for low frequency components of the spectra is necessary to address periodic seasonal fluctuations and long-term drifting. Because no low frequency drift occurred in tested Bruins Instruments analyzers, removal of low frequency coefficients was not attempted.

The dvantage of wavelets multilevel decomposition as a preprocessing method for calibration development is its flexibility. Besides ability to remove high frequency noise from spectra, wavelet decomposition has potential to selectively remove low frequency drift from the spectra.

REFERENCES

[1] Foo-tim Chau and Alexander Kai-man Leung, "Applications of Wavelets in Spectroscopic Studies", Wavelets in Chemistry Edited by B. Waltzak, ©2000 Elsevier Science B.V., p.253

[2] Hu-Wei Tan and Steven D. Brown, "Wavelet analysis applied t o removing non-constant, varying spectroscopic background in multivariate calibration", Journal of Chemometrics, 16, © 2002 Wiley InterScience, p. 228

[3] I. Murray, "Scattered information: philosophy and practice of near infrared spectroscopy", Near Infrared Spectroscopy: Proceeding of the 11th International Conference, Edited by A.M.C Davies and A. Garrido-Varo, p.11

[4] Beata Walczak and Jan P. Radomski, "Wavelet Bases for IR Library Compression, Searching and Reconstruction", Wavelets in Chemistry Edited by B. Waltzak, ©2000 Elsevier Science B.V., p.297

[5] Pekka Teppola and Pentti Minkinen, "Wavelet-PLS regression models for both exploratory data analysis and process monitoring", Journal of Chemometrics, 14, © 2000 Wiley InterScience, p. 383

[6] Robert Dzupin, Sylvie A. Rousell, Glen R. Rippke, and CharlesR. Hurburgh, Jr., NIR_GrainNET: Internet-enabled software for real-time prediction using NIR whole-grain analyzers, PITTCON, 2002

[7] R. James Berry and Yukihiro Ozaki, "Comparison of Wavelets and Smoothing for Denoising Spectra for Two-Dimensional Correlation Spectroscopy", Applied Spectroscopy, © 2002 Society for Applied Spectroscopy, Volume 56, Number 11, 2002, p 1463

[8] Frederic Despagne, Beata Walczak, and Desire-Luc Massart, "Transfer of Calibrations of Near-Infrared Spectra Using Neural Networks", Applied Spectroscopy, © 1998 Society for Applied Spectroscopy, Volume 52, Number 5, 1998, p 733

[9] P.C. Williams, "Implementation of Near-Infrared Technology", Near-Infrared Technology in the Agricultural and Food Industries", Edited by Phil Williams and Karl Norris, ©2001 by the American Association of Cereal Chemists, Inc., p. 164

GENERAL CONCLUSIONS

The GrainNet software was designed to address the incorporation of highly sophisticated mathematical algorithms into the computer instrumentation used to extract information from raw spectral data. GrainNet software connects the SQL Server database with mathematical algorithms developed in a MATLAB® computational environment, and provides real-time predictions of near-infrared analyzes.

The "ultimate" system, where calibration is based on samples supplied by diverse clients to a host laboratory, and is used to predict results upon receipt of spectra by e-mail, using the local or Artificial Neural Networks (ANN) models, was proposed by Phil Williams. GrainNet software is extending the idea of the "ultimate" system to real-time and the possibility of improving accuracy of prediction by center averaging the results of several models or choosing models based on sample properties.

Several standardization transfer approaches were studied in this project. Direct standardization was a sufficient method to successfully transfer spectral data from Bruins Instruments analyzers to Foss Infratec 1241 calibration model for the tested population of analyzers available in the Grain Quality Laboratory.

The Multiplicative Scatter Correction (MSC) pretreatment, Piecewise Direct Standardization (PDS), and Finite Impulse Response (FIR) were also tested and evaluated for future use with the GrainNet software. The new algorithm of wavelets multilevel decomposition in combination with the ANN regression model for spectral noise removal was developed and studied. The new calibration model successfully removed noise from the spectra of Bruins Instruments and produced reliable protein and oil content predictions.

Use of calibration chemometrics models in NIR spectroscopy is very often limited by prediction software. A calibration that can predict with 100% precision and accuracy is useless if spectrometer cannot use it. A similar situation is true for standardization. The main reason using of simple slope and bias correction for standardization is often that this is only technique that prediction software can implement. The GrainNet software is designed to solve this problem and bridge the gap between hardware of the spectrometers and available customized chemometrics routines. The GrainNet software is developed to substitute vendor specific prediction software in order to provide a "research friendly" environment for development and implementation of new ideas for calibration and standardization.

APPENDIX A. SCREENSHOTS OF SOFTWARE USED IN THE PROJECT

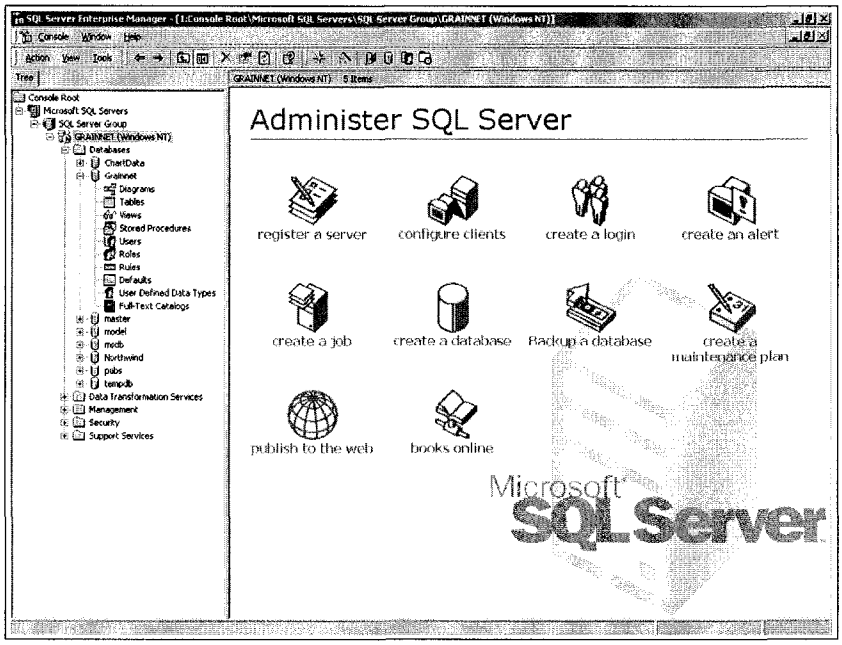

Figure A1. SQL Server database used in GrainNet software.

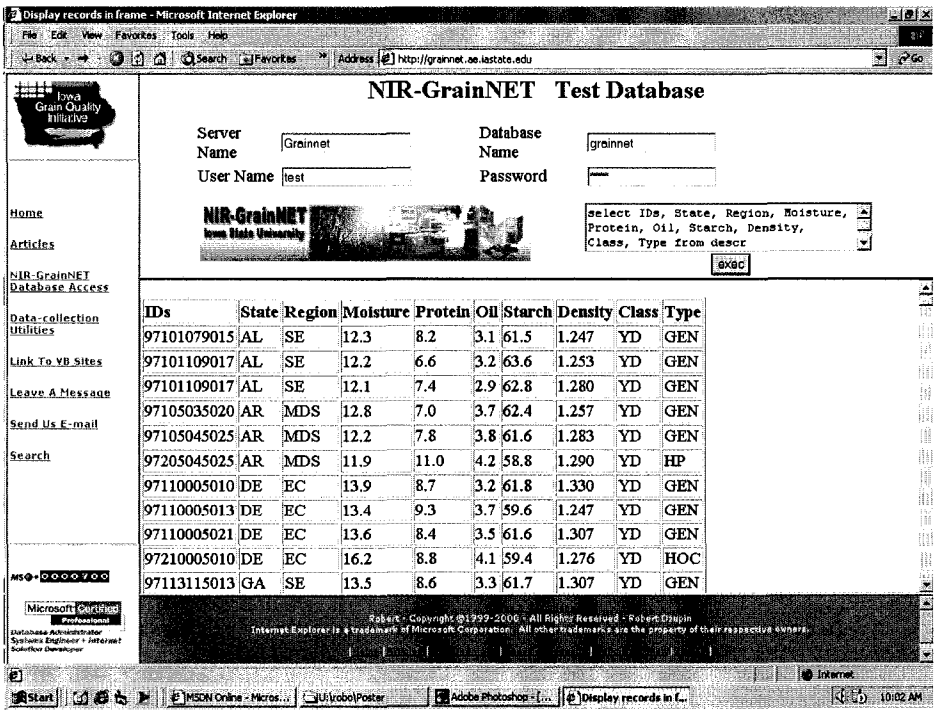

Figure A2. GrainNet Internet Information Server that serves to access data in SQL Server

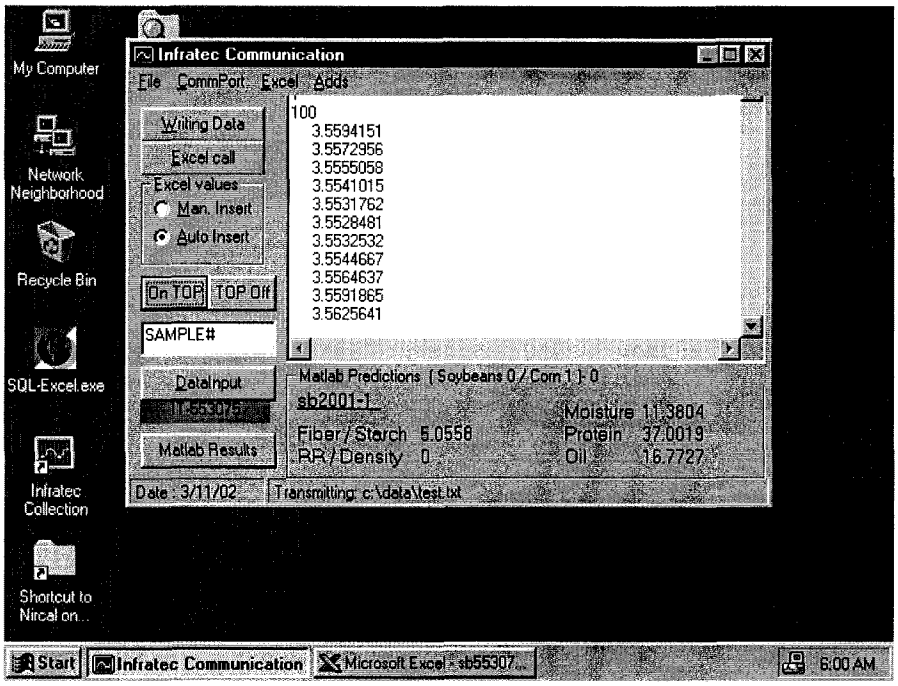

Figure A3. GrainNet client software for Foss Infratec 1225 and 1229. Client is using serial port to communicate with Foss analyzer. Spectra are sent to SQL server over Internet. Predictions are calculated on the server and in real-time send back to the client.

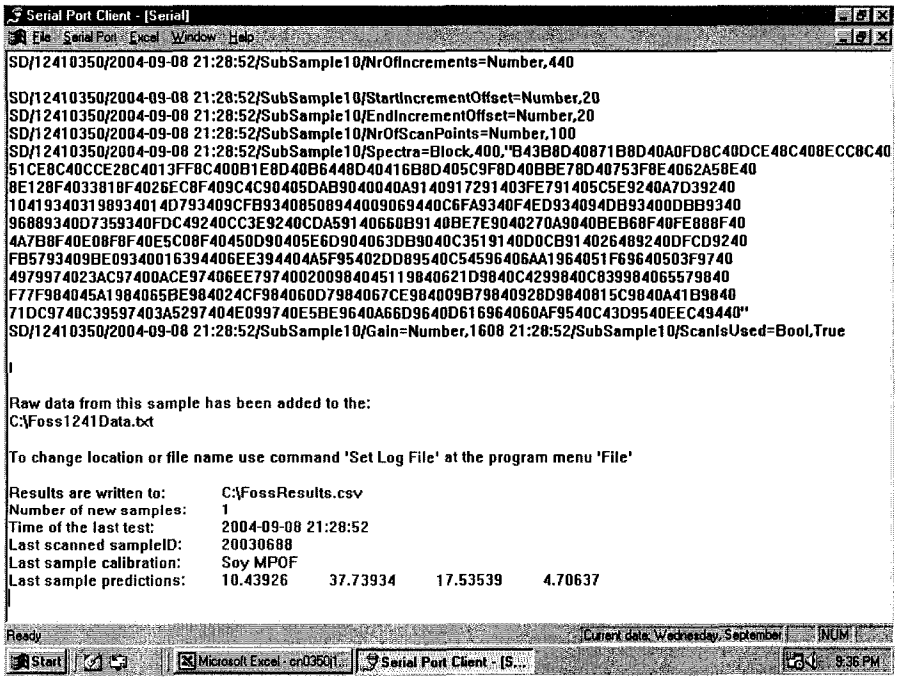

Figure A4. GrainNet client software for Foss Infratec 1241. Displayed results are directly from analyzer. Client software is used for decoding data from IEEE 754 float format to ASCII and to communicate with MS Excel.

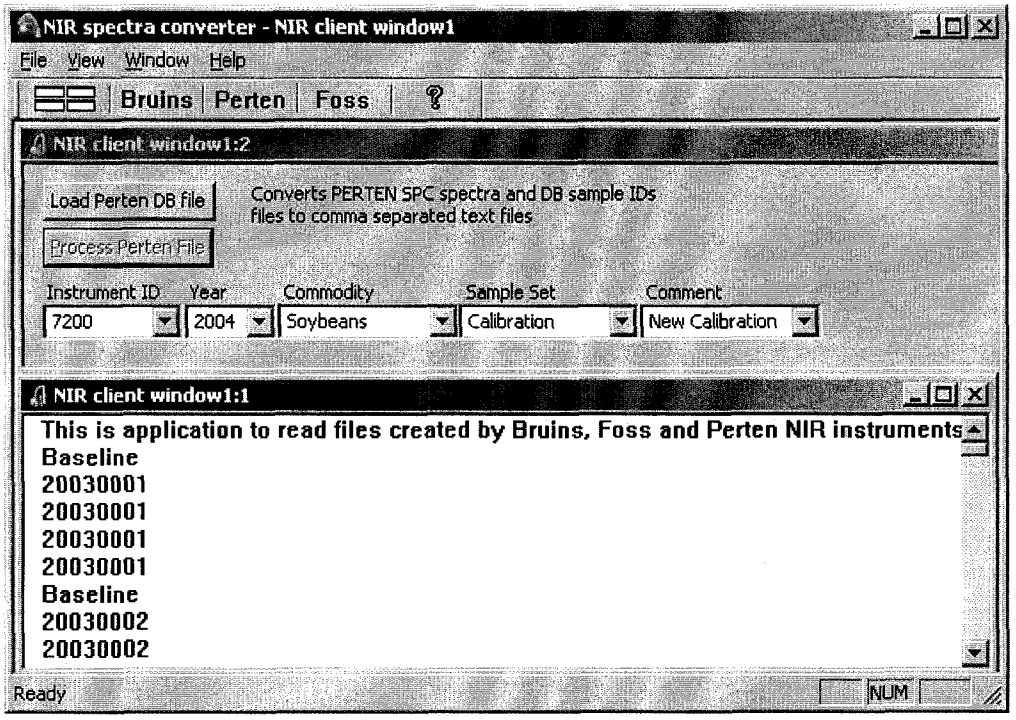

Figure A5. GrainNet client software for Foss 1241, Bruins Instruments, and Perten 7000 and 7200. Software is importing and exporting data between MS SQL server database, Matlab®, and mentioned analyzers.

APPENDIX B. LISTING OF THE MATLAB SCRIPT USED IN THE PROJECT

```
clear all; close all; disp('Script started');
SelectedCommodity=2 ;                % Protein=1, Oil=2, Starch=3 Moisture=4
FossData=1  ;                        % Enable FOSS Section.
BruinsData=0;                        % Enable Bruins Section.
if BruinsData==1
FossData=0 ;                         % Disable processing of Foss inputs
FOSSrangeOnly=1;                     % Reduce bruins data to range 850-1048
BruinsPseudoStandardisation=1;       % Pseudo-standardisation use all samples
                                     % for 2nd and 3rd Bruins instrument
end
FossBruins=8;          % if value is 8 than do also calculations for Bruins instruments
                       % using Foss calibration
%************************** STANDARDIZATION ********************
Run_Wavelets=1;                                                          %*
SkipPCAforNN=1;        % When Wavelets are used instead of the PCAs      %*
level=4;               % how many compress loops                        %*
w='db2';               %name of wavelet funct.                         %*
useWaveletDetails=0;                                                     %*
Run_MSCorr_Standardization=0;                                            %*
Run_FIR_Standardization=0;                                               %*
FIRwinSize= 29;        %FOSS best window size is 29                    %*
% FIRwinSize= 35;      %BRUINS best window size is 35                  %*
Run_PDS_Standardization=0;                                               %*
PDS_window=0 ;                                                           %*
Run_SlopeBias=0;                                                         %*
%************************************************************************

%*************** NEURAL NETWORK ACTIVATION ******************
RecalculateNN=1  ;     % If equal 1 then overwrite loaded NN model       %*
DisplayNN=1  ;         % Show NN results (R2, RPD, STD ...)              %*
%************************************************************************

%SELECT BRUINS models
% load lastBruinsNN;
% load trainedNN_BruinsMoisture01.mat;
% load trainedNN_BruinsProtein01.mat;
% load trainedNN_BruinsOil01.mat;
% load trainedFIR_NN_BruinsMoisture01.mat;
% load trainedFIR_NN_BruinsProtein01.mat;
% load trainedFIR_NN_BruinsOil01.mat;
% load trainedFIR_NN_BruinsStarch01.mat;
```

```
% load trainedMSC_NN_BruinsMoisture01.mat;
% load trainedMSC_NN_BruinsProtein01.mat;
% load trainedMSC_NN_BruinsOil01.mat;
% load trainedMSC_NN_BruinsStarch01.mat;
%load WaveletMoisture_db6_5.mat;
%load WaveletProtein_db6_5.mat;
%load WaveletOil_db6_5.mat;



%********* SELECT FOSS models  **************
%       load trainedNN_FOSSmoisture01.mat;
%       load trainedNN_FOSSprotein01.mat;
        load trainedNN_FOSSoil01.mat;
%       load trainedFIR_NN_FOSSmoisture01.mat;
%       load trainedFIR_NN_FOSSprotein01.mat;
%       load trainedFIR_NN_FOSSoil01.mat;
%       load trainedFIR_NN_FOSSstarch01.mat;
%       load trainedMSC_NN_FOSSmoisture01.mat;
%       load trainedMSC_NN_FOSSprotein01.mat;
%       load trainedMSC_NN_FOSSoil01.mat;
%       load trainedMSC_NN_FOSSstarch01.mat;

% roboPCA_Variance=0.0001;              % Use ONLY for FIR standardisation
                                        % (for Bruins 0.003) (for FOSS 0.000
roboPCA_Variance=1e-8;                  % use for MScorr and RAW data
% roboPCA_Variance=0.003;               % Use for Bruins data FIR
roboNumberOfNodes=3;
roboNumberOfEpochs=1200;
disp(sprintf('ANN parameters used: pca=%1.0e nodes=%1i epochs=%2i',
roboPCA_Variance, roboNumberOfNodes,roboNumberOfEpochs));
%*****************************************************************



%********************** PLS MODEL *******************************
CalculatePLS=1;                 % Enable PLS calculation
SelectPLStoCheck=3 ;            % Select PLS model
%BRUINS CALCULATIONS                            FOSS CALCULATIONS
% 1= instrumentCalibration ='Bruins 6110 ';     'FOSS 1241  ALL calib. scans';
% 2= instrumentValidation     ='Bruins 6110 '    'FOSS 1241   standardisation';
% 3= instrumentStandardisation ='Bruins 310002' 'FOSS 0065   standardisation';
% 4= instrumentStandardisation2 ='Bruins 6118 ' 'FOSS 3108   standardisation';
% 5= instrumentStandardisation3 ='Bruins 6175 ' 'FOSS 553792 standardisation';
%******************** END OF SETTTING SECTION  ******************
```

```
%********** LOAD FOSS SPECTRA and REFERENCE DATA **********
if FossData==1
%One daily check sample with 30 repeats
Corn1241robustness=dlmread('FOSS_robustness.csv',',');
Corn1241Nov2004Moisture=dlmread('CornFOSSmoisture.csv',',');
Corn1241Nov2004Protein=dlmread('CornFOSSprotein.csv',',');
Corn1241Nov2004Oil=dlmread('CornFOSSoil.csv',',');
Corn1241Nov2004Starch=dlmread('CornFOSSstarch.csv',',');

%Retrieve FOSS ROBUSTNESS data for calibration and standardisation by using
1241
FOSS1241_Robustness=Corn1241robustness(31:60,:);
FOSS553792_Robustness=Corn1241robustness(61:90,:);
FOSS3108_Robustness=Corn1241robustness(91:118,:);
FOSS0065_Robustness=Corn1241robustness(119:149,:);

%Retrieve FOSS MOISTURE data for calibration and standardisation for moisture
FOSS1241_Moisture_All=Corn1241Nov2004Moisture(1:464,:);
FOSS1241_Moisture_std=Corn1241Nov2004Moisture(465:484,:);
FOSS553792_Moisture_std=Corn1241Nov2004Moisture(485:504,:);
FOSS3108_Moisture_std=Corn1241Nov2004Moisture(505:523,:);
FOSS0065_Moisture_std=Corn1241Nov2004Moisture(524:543,:);

%Retrieve FOSS PROTEIN data for calibration and standardisation for protein
FOSS1241_Protein_cal=Corn1241Nov2004Protein(1:455,:);
FOSS1241_Protein_std=Corn1241Nov2004Protein(456:475,:);
FOSS553792_Protein_std=Corn1241Nov2004Protein(476:495,:);
FOSS3108_Protein_std=Corn1241Nov2004Protein(496:514,:);
FOSS0065_Protein_std=Corn1241Nov2004Protein(515:534,:);
    %FOSS1241_Protein_std=Corn1241Nov2004Protein(394:464,:);

    %Retrieve FOSS OIL data for calibration and standardisation for oil
    FOSS1241_Oil_cal=Corn1241Nov2004Oil(1:464,:);
    FOSS1241_Oil_std=Corn1241Nov2004Oil(465:484,:);
    FOSS553792_Oil_std=Corn1241Nov2004Oil(485:504,:);
    FOSS3108_Oil_std=Corn1241Nov2004Oil(505:523,:);
    FOSS0065_Oil_std=Corn1241Nov2004Oil(524:543,:);

    %Retrieve FOSS STARCH data for calibration and standardisation starch
    FOSS1241_Starch_cal=Corn1241Nov2004Starch(1:118,:);
    FOSS1241_Starch_std=Corn1241Nov2004Starch(119:127,:);
    FOSS553792_Starch_std=Corn1241Nov2004Starch(128:136,:);
    FOSS3108_Starch_std=Corn1241Nov2004Starch(137:144,:);
    FOSS0065_Starch_std=Corn1241Nov2004Starch(145:153,:);
```

```
% Retrieve data for Optical Standardisation (using stdsslct function I found
% indexes of 5 samples that I will use for optical standardisation)
% subsamples=[150 181 232 362 387];
% Samples selected: 20000056, 20000162, 20000670, 20020351, 20030024
% FOSS1241_5=FOSS1241_Protein_cal(subsamples,:);

if SelectedCommodity==1
    calibration=FOSS1241_Protein_cal;
    validation=FOSS1241_Protein_std;
    standardisation=FOSS0065_Protein_std;
    standardisation2=FOSS3108_Protein_std;
    standardisation3=FOSS553792_Protein_std;
    robustness1241=FOSS1241_Robustness;
    robustness0065=FOSS553792_Robustness;
    robustness3108=FOSS3108_Robustness;
    robustness553792=FOSS0065_Robustness;
end

if SelectedCommodity==2
    calibration=FOSS1241_Oil_cal;
    validation=FOSS1241_Oil_std;
    standardisation=FOSS0065_Oil_std;
    standardisation2=FOSS3108_Oil_std;
    standardisation3=FOSS553792_Oil_std;
    robustness1241=FOSS1241_Robustness;
    robustness0065=FOSS553792_Robustness;
    robustness3108=FOSS3108_Robustness;
    robustness553792=FOSS0065_Robustness;
end

if SelectedCommodity==3
    calibration=FOSS1241_Starch_cal;
    validation=FOSS1241_Starch_std;
    standardisation=FOSS0065_Starch_std;
    standardisation2=FOSS3108_Starch_std;
    standardisation3=FOSS553792_Starch_std;
    robustness1241=FOSS1241_Robustness;
    robustness0065=FOSS553792_Robustness;
    robustness3108=FOSS3108_Robustness;
    robustness553792=FOSS0065_Robustness;
end

if SelectedCommodity==4
    calibration=FOSS1241_Moisture_cal;
    validation=FOSS1241_Moisture_std;
```

```
154             standardisation=FOSS0065_Moisture_std;
155             standardisation2=FOSS3108_Moisture_std;
156             standardisation3=FOSS553792_Moisture_std;
157             robustness1241=FOSS1241_Robustness;
158             robustness0065=FOSS553792_Robustness;
159             robustness3108=FOSS3108_Robustness;
160             robustness553792=FOSS0065_Robustness;
161         end


162         %************** Assign scan and proximate values ******************
163         scans=calibration(:,6:end);
164         proximate=calibration(:,5);

165         scansValidation=validation(:,6:end);
166         proximateValidation=validation(:,5);

167         scansStandardisation=standardisation(:,6:end);
168         proximateStandardisation=standardisation(:,5);

169         scansStandardisation2=standardisation2(:,6:end);
170         proximateStandardisation2=standardisation2(:,5);

171         scansStandardisation3=standardisation3(:,6:end);
172         proximateStandardisation3=standardisation3(:,5);

173         scansRobustness1=robustness1241(:,6:end);
174         proximateRobustness1=robustness1241(:,5);

175         scansRobustness2=robustness0065(:,6:end);
176         proximateRobustness2=robustness0065(:,5);

177         scansRobustness3=robustness3108(:,6:end);
178         proximateRobustness3=robustness3108(:,5);

179         scansRobustness4=robustness553792(:,6:end);
180         proximateRobustness4=robustness553792(:,5);


181         %********************** Save FOSS to separate File ****************
182         scansFOSS=scans;
183         proximateFOSS=proximate;
184         %*************************************************************
```

```
185     %**************** FOSS model using Bruins spectra ****************
186     if FossBruins==8
187     %FIRST READ ORIGINAL SPC and TEXT files and concate them together
                1. spcreadr('Corn310002Nov2004c.spc');
                   Corn310002Nov2004c=ans;
                2. spcreadr('Corn6110Nov2004c.spc'); Corn6110Nov2004c=ans;
                3. spcreadr('Corn6118Nov2004c.spc'); Corn6118Nov2004c=ans;
                4. spcreadr('Corn6175Nov2004std.spc');
                   Corn6175Nov2004std=ans;
188     % Load tab delimited text files
                1. Corn6110Nov2004=dlmread('Corn6110Nov2004c.txt','\t');
                2. %Reading text file with 5 columns:
                   SampleID+Moisture+Protein+Oil+Starch
                3. Corn6118Nov2004=dlmread('Corn6118Nov2004c.txt','\t');
                4. Corn310002Nov2004=dlmread('Corn31002Nov2004c.txt','\t');
                5. Corn6175Nov2004=dlmread('Corn6175Nov2004std.txt','\t');
189     % Merge data from txt and spc files together
190     Corn310002Nov2004_All=[Corn310002Nov2004,Corn310002Nov2004c];
                1. %Data will now be complete ID+proximate+scans
                2. Corn6110Nov2004_All=[Corn6110Nov2004,Corn6110Nov20
                   04c];
                3. Corn6118Nov2004_All=[Corn6118Nov2004,Corn6118Nov20
                   04c];
                4. Corn6175Nov2004_All=[Corn6175Nov2004,Corn6175Nov20
                   04std];
191     %********************** WARNING ***************************
                1. %Change values if different number of samples is used
                2. %Extract Standardization from calibration data
                3. Corn310002Nov2004_std=Corn310002Nov2004_All(363:393,
                   :);%
                4. Corn6110Nov2004_std=Corn6110Nov2004_All(363:393,:);
                5. Corn6118Nov2004_std=Corn6118Nov2004_All(365:395,:);
                6. Corn6175Nov2004_std=Corn6175Nov2004_All;
192     Corn6110Nov2004_std=sortrows(Corn6110Nov2004_std);
193     Corn6118Nov2004_std=sortrows(Corn6118Nov2004_std);
194     Corn310002Nov2004_std=sortrows(Corn310002Nov2004_std);
195     Corn6175Nov2004_std=sortrows(Corn6175Nov2004_std);

196     %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
197     Optical1241=FOSS1241_Oil_std;
198     % original data have 746 points (IDs, moist,prot,oil,starch ...741 scans)
199     % I need FOSS infratec range scan data
200     Optical1241=sortrows(Optical1241);
201     master=Optical1241;          % optical scans values for standardisation set
                                     % from calibration instrument
```

```
202     spcreadr('Corn6175Nov2004std.spc'); Corn6175Nov2004std=ans;
203     Corn6175Nov2004=dlmread('Corn6175Nov2004std.txt','\t');
204     Corn6175Nov2004_All=[Corn6175Nov2004,Corn6175Nov2004std];
205     Corn6175Nov2004_All=sortrows(Corn6175Nov2004_All);

206     % SET Different Bruins Instruments to be used with FOSS standardisation
207     % Corn6175Nov2004_All=Corn6110Nov2004_std;
208     % REPLACE 6175 with 6110
209     % Corn6175Nov2004_All=Corn6110Nov2004_std;
210     % REPLACE 6175 with 6118
211     % Corn6175Nov2004_All=Corn6110Nov2004_std;
212     % REPLACE 6175 with 310002

213     slave=Corn6175Nov2004_All(:,244:639) ;
214     [std_matrix, std_vector]=stdgen(master,slave,PDS_window);

215     scansStandardisation=stdize(slave,std_matrix,std_vector);
216     if SelectedCommodity==4  %moisture
                proximateStandardisation=Corn6175Nov2004_All(:,2);
217     end
218     if SelectedCommodity==1  %protein
                clear proximateStandardisation;
                proximateStandardisation=Corn6175Nov2004_All(:,3);
219     end
220     if SelectedCommodity==2  %oil
                proximateStandardisation=Corn6175Nov2004_All(:,4);
221     end
222     if SelectedCommodity==3  %starch
                proximateStandardisation=Corn6175Nov2004_All(:,5);
223     end
224     %*************** END of FOSS using Bruins slave ******************
.
.
.
.

1488    if (FossData==1 )
1489    %_____________ SIMULATE NN and print results _____________
            1. if SelectedCommodity==1  disp('Values below are calculated
               for PROTEIN'); end
            2. if SelectedCommodity==2  disp('Values below are calculated
               for OIL')  ;end
            3. if SelectedCommodity==3  disp('Values below are calculated
               for STARCH')  ;end
```

```
                4. if SelectedCommodity==4  disp('Values below are calculated
                   for MOISTURE') ;end
1490    %Display results of prediction for VALIDATION set
                1. figure ('Name','Validation');%VALIDATION
                2. pnewn=trastd(scansValidation',meanp,stdp);
                3. pnewtrans=trapca(pnewn,transMat);
1491    if SkipPCAforNN==1
            pnewtrans=pnewn;
1492    end
1493    a=sim(net,pnewtrans);
1494    [m,b,r]=postreg(a,proximateValidation');
1495    R2_Validation=r*r;
1496    SEP=std(a-proximateValidation');
1497    RPD=std(proximateValidation')/SEP;
1498    disp(sprintf('Instrument Validation:\t\t\t R^2=%0.5g  \t RPD=%0.5g \t
                    SEP=%0.5g   ',R2_Validation,RPD,SEP));
1499    %Display results of prediction for CALIBRATION set
1500    %figure ('Name','Calibration');%CALIBRATION
1501    pnewn=trastd(scans',meanp,stdp);
1502    pnewtrans=trapca(pnewn,transMat);
1503    if SkipPCAforNN==1
            pnewtrans=pnewn;
1504    end
1505    a=sim(net,pnewtrans);
1506    [m,b,r]=postreg(a,proximate');
1507    R2_Calibration=r*r;
1508    SEP=std(a-proximate');
1509    RPD=std(proximate')/SEP;
1510    disp(sprintf('Instrument Calibration:\t\t\t R^2=%0.5g \t RPD=%0.5g \t
1511    SEP=%0.5g',R2_Calibration,RPD,SEP));

1512    %_____Load standardisation samples to ANN_____

1513    %figure ('Name','Standardisation');%STANDARDISATION

1514    pnewn=trastd(scansStandardisation',meanp,stdp);
1515    pnewtrans=trapca(pnewn,transMat);
1516    if SkipPCAforNN==1
            pnewtrans=pnewn;
1517    end
1518    a=sim(net,pnewtrans);
1519    [m,b,r]=postreg(a,proximateStandardisation');
1520    R2_Standardisation_0065=r*r;
1521    SEP=std(a-proximateStandardisation');
1522    RPD=std(proximateStandardisation')/SEP;
```

```
1523    disp(sprintf('Instrument 0065:\t\t\t\t R^2=%0.5g \t RPD=%0.5g \t SEP=%0.5g
  ',R2_Standardisation_0065,RPD,SEP));
1524    %_____Load standardisation samples to ANN_____
1525    %figure ('Name','Standardisation2');%STANDARDISATION
1526    pnewn=trastd(scansStandardisation2',meanp,stdp);
1527    pnewtrans=trapca(pnewn,transMat);
1528    if SkipPCAforNN==1
            pnewtrans=pnewn;
1529    end
                1. a=sim(net,pnewtrans);
                2. [m,b,r]=postreg(a,proximateStandardisation2');
                3. R2_Standardisation2_3108=r*r;
                4. SEP=std(a-proximateStandardisation2');
                5. RPD=std(proximateStandardisation2')/SEP;
                6. disp(sprintf('Instrument 3108:\t\t\t\t R^2=%0.5g \t RPD=%0.5g
                   \t SEP=%0.5g   ',R2_Standardisation2_3108,RPD,SEP));


                7. %_____Load standardisation samples to ANN_____
                8. %figure ('Name','Standardisation3');%STANDARDISATION
                9. pnewn=trastd(scansStandardisation3',meanp,stdp);
                10. pnewtrans=trapca(pnewn,transMat);
1530    if SkipPCAforNN==1
    a.  pnewtrans=pnewn;
1531    end
                1. a=sim(net,pnewtrans);
1532    a(1)=[];
1533    proximateStandardisation3(1)=[];
1534    [m,b,r]=postreg(a,proximateStandardisation3') ;
1535    R2_Standardisation3_553792=r*r;
1536    SEP=std(a-proximateStandardisation3');
1537    RPD=std(proximateStandardisation3')/SEP;
1538    disp(sprintf('Instrument 553792:\t\t\t\t R^2=%0.5g \t RPD=%0.5g \t
  SEP=%0.5g ',R2_Standardisation3_553792,RPD,SEP));

1539    % ROBUSTNESS1 calculation
                1. %figure ('Name','Standardisation3');%STANDARDISATION
                2. pnewn=trastd(scansRobustness1',meanp,stdp);
                3. pnewtrans=trapca(pnewn,transMat);
1540    if SkipPCAforNN==1
            pnewtrans=pnewn;
1541    end
```

```
1759        %INFORMATION ABOUT TESTS
1760        if (FossData==1)  disp('FOSS');   end
1761        if (BruinsData==1)  disp('BRUINS');  end
1762        if (SelectedCommodity==1)    disp('Protein'); end
1763        if (SelectedCommodity==2)    disp('Oil'); end
1764        if (SelectedCommodity==3)    disp('Starch'); end
1765        if (SelectedCommodity==4)    disp('Moisture'); end
1766        if (Run_Wavelets==1)
1767        asize=size(pnewn);
1768        disp(sprintf('WAVELETS type=%s \t level=%i \t Inputs=%0.5g
    ',w,level,asize(1,1)));
1769        end
1770        if (Run_MSCorr_Standardization==1)  disp('MSCorr'); end
1771        if (Run_FIR_Standardization==1)    disp('FIR'); end
1772        if (Run_PDS_Standardization==1)    disp('PDS standardisation  ');
    disp(PDS_window);
1773        end
1774        %********************** END OF SCRIPT **********************
```

APPENDIX C. MATLAB CODE OUTPUTS

ANN parameters used: pca	1e-008 nodes	3 epochs	1200			
Values below are calculated for MOISTURE						
Bruins 6110 :	R^2	0.99171	RPD	10.983	SEP	0.19710
Bruins 6110 :	R^2	0.99271	RPD	11.539	SEP	0.11199
Bruins 310002:	R^2	0.98711	RPD	7.8903	SEP	0.16378
Bruins 6118 :	R^2	0.96983	RPD	5.6818	SEP	0.22709
Bruins 6175 :	R^2	0.99023	RPD	10.094	SEP	0.12619
ANN parameters used:	pca	1.00E-08	nodes	3	epochs	1200
Instrument 6110 Robustness:					SEP	0.14118
Instrument 6118 Robustness:					SEP	0.71676
Instrument 6175 Robustness:					SEP	0.16473
Instrument 31002 Robustness:					SEP	0.21635
Bias Corrected Instrument 6118:	R^2	0.96983	RPD	5.7463	SEP	0.22454
Bias Corrected Instrument 310002:	R^2	0.98711	RPD	5.9801	SEP	0.21609

Figure C1. Matlab Script Output. Moisture predictions for 'Bruins No Standardization'

ANN parameters used: pca	1e-008 nodes	3 epochs	1200			
Values below are calculated for PROTEIN						
Bruins 6110 :	R^2	0.96649	RPD	5.4625	SEP	0.25544
Bruins 6110 :	R^2	0.98348	RPD	7.771	SEP	0.26298
Bruins 310002:	R^2	0.96013	RPD	4.3893	SEP	0.46558
Bruins 6118 :	R^2	0.98259	RPD	7.0977	SEP	0.29227
Bruins 6175 :	R^2	0.97881	RPD	6.8635	SEP	0.28983
ANN parameters used:	pca	1.00E-08	nodes	3	epochs	1200
Instrument 6110 Robustness:					SEP	0.08355
Instrument 6118 Robustness:					SEP	0.28519
Instrument 6175 Robustness:					SEP	0.1369
Instrument 31002 Robustness:					SEP	0.32683
Bias Corrected Instrument 6118:	R^2	0.98259	RPD	7.442	SEP	0.27875
Bias Corrected Instrument 310002:	R^2	0.96013	RPD	2.987	SEP	0.68416

Figure C2. Matlab Script Output. Protein predictions for 'Bruins No Standardization'

ANN parameters used: pca	1e-008 nodes	3 epochs	1200			
Values below are calculated for OIL						
Bruins 6110 :	R^2	0.95726	RPD	4.837	SEP	0.2572
Bruins 6110 :	R^2	0.96562	RPD	5.1875	SEP	0.23003
Bruins 310002:	R^2	0.7626	RPD	2.0422	SEP	0.58432
Bruins 6118 :	R^2	0.97183	RPD	5.9234	SEP	0.20385
Bruins 6175 :	R^2	0.97225	RPD	6.0032	SEP	0.19441
ANN parameters used:	pca	1.00E-08	nodes	3	epochs	1200
Instrument 6110 Robustness:					SEP	0.13533
Instrument 6118 Robustness:					SEP	0.19412
Instrument 6175 Robustness:					SEP	0.16172
Instrument 31002 Robustness:					SEP	0.55137
Bias Corrected Instrument 6118:	R^2	0.97183	RPD	5.8366	SEP	0.20688
Bias Corrected Instrument 310002:	R^2	0.7626	RPD	2.0333	SEP	0.58687

Figure C3. Matlab Script Output. Oil predictions for 'Bruins No Standardization'

ANN parameters used: pca	1e-008 nodes	3 epochs	1200			
Values below are calculated for MOISTURE						
Bruins 6110 :	R^2	0.98949	RPD	9.7539	SEP	0.2225
Bruins 6110 :	R^2	0.98774	RPD	8.9625	SEP	0.14418
Bruins 310002:	R^2	0.9756	RPD	6.3752	SEP	0.2027
Bruins 6118 :	R^2	0.98261	RPD	7.4659	SEP	0.17282
Bruins 6175 :	R^2	0.98457	RPD	7.9764	SEP	0.1597
ANN parameters used:	pca	1.00E-08	nodes	3	Epochs	1200
Instrument 6110 Robustness:					SEP	0.171
Instrument 6118 Robustness:					SEP	0.17182
Instrument 6175 Robustness:					SEP	0.13531
Instrument 31002 Robustness:					SEP	0.28212
Bias Corrected Instrument 6118:	R^2	0.98261	RPD	1.2262	SEP	1.0523
Bias Corrected Instrument 310002:	R^2	0.9756	RPD	1.1026	SEP	1.172

Figure C4. Matlab Script Output. Moisture predictions for 'Bruins with FIR Standardization'

ANN parameters used: pca	1e-008 nodes	3 epochs	1200			
Values below are calculated for PROTEIN						
Bruins 6110 :	R^2	0.9574	RPD	4.8445	SEP	0.28802
Bruins 6110 :	R^2	0.98066	RPD	7.0973	SEP	0.28794
Bruins 310002:	R^2	0.94689	RPD	3.9083	SEP	0.52289
Bruins 6118 :	R^2	0.98095	RPD	7.1245	SEP	0.29117
Bruins 6175 :	R^2	0.97333	RPD	6.121	SEP	0.32498
ANN parameters used:	pca	1.00E-08	nodes	3	epochs	1200
Instrument 6110 Robustness:					SEP	0.21491
Instrument 6118 Robustness:					SEP	0.56919
Instrument 6175 Robustness:					SEP	0.37475
Instrument 31002 Robustness:					SEP	0.68486
Bias Corrected Instrument 6118:	R^2	0.98095	RPD	1.6034	SEP	1.2938
Bias Corrected Instrument 310002:	R^2	0.94689	RPD	0.68454	SEP	2.9853

Figure C5. Matlab Script Output. Protein predictions for 'Bruins with FIR Standardization'

ANN parameters used: pca	1e-008 nodes	3 epochs	1200			
Values below are calculated for OIL						
Bruins 6110 :	R^2	0.95639	RPD	4.7881	SEP	0.25983
Bruins 6110 :	R^2	0.96112	RPD	5.0124	SEP	0.23806
Bruins 310002:	R^2	0.94594	RPD	4.2994	SEP	0.27754
Bruins 6118 :	R^2	0.95328	RPD	4.4067	SEP	0.27401
Bruins 6175 :	R^2	0.95696	RPD	4.6344	SEP	0.25183
ANN parameters used:	pca	1.00E-08	nodes	3	epochs	1200
Instrument 6110 Robustness:					SEP	0.094959
Instrument 6118 Robustness:					SEP	0.29003
Instrument 6175 Robustness:					SEP	0.1577
Instrument 31002 Robustness:					SEP	0.46851
Bias Corrected Instrument 6118:	R^2	0.95328	RPD	0.53774	SEP	2.2455
Bias Corrected Instrument 310002:	R^2	0.94594	RPD	0.75254	SEP	1.5857

Figure C6. Matlab Script Output. Oil predictions for 'Bruins with FIR Standardization'

ANN parameters used: pca		1e-008 nodes	3 epochs	1200			
Values below are calculated for MOISTURE							
Bruins 6110 :	R^2		0.99206	RPD	11.221	SEP	0.19342
Bruins 6110 :	R^2		0.99362	RPD	12.435	SEP	0.10392
Bruins 310002:	R^2		0.97894	RPD	6.867	SEP	0.31613
Bruins 6118 :	R^2		0.98756	RPD	8.9158	SEP	0.24245
Bruins 6175 :	R^2		0.98957	RPD	9.7884	SEP	0.13013
ANN parameters used:	pca		1.00E-08	nodes	3	epochs	1200
Instrument 6110 Robustness:						SEP	0.14109
Instrument 6118 Robustness:						SEP	0.24678
Instrument 6175 Robustness:						SEP	0.13201
Instrument 31002 Robustness:						SEP	0.2342
Bias Corrected Instrument 6118:	R^2		0.98756	RPD	8.8035	SEP	0.24554
Bias Corrected Instrument 310002:	R^2		0.97894	RPD	6.2491	SEP	0.34739

Figure C7. Matlab Script Output. Moisture predictions for 'Bruins with MSC pretreatment'

ANN parameters used: pca	1e-008 nodes	3 epochs	1200			
Values below are calculated for PROTEIN						
Bruins 6110 :	R^2	0.9644	RPD	5.3001	SEP	0.26326
Bruins 6110 :	R^2	0.98254	RPD	7.1706	SEP	0.28499
Bruins 310002:	R^2	0.91181	RPD	3.3649	SEP	0.41377
Bruins 6118 :	R^2	0.95576	RPD	4.6751	SEP	0.29788
Bruins 6175 :	R^2	0.98017	RPD	6.9357	SEP	0.28681
ANN parameters used:	pca	1.00E-08	nodes	3	epochs	1200
Instrument 6110 Robustness:					SEP	0.1007
Instrument 6118 Robustness:					SEP	0.37472
Instrument 6175 Robustness:					SEP	0.15841
Instrument 31002 Robustness:					SEP	0.33012
Bias Corrected Instrument 6118:	R^2	0.95576	RPD	4.7521	SEP	0.29306
Bias Corrected Instrument 310002:	R^2	0.91181	RPD	2.8835	SEP	0.48286

Figure C8. Matlab Script Output. Protein predictions for 'Bruins with MSC pretreatment'

ANN parameters used: pca		1e-008 nodes	3 epochs		1200	
Values below are calculated for OIL						
Bruins 6110 :	R^2	0.95537	RPD	4.733	SEP	0.26285
Bruins 6110 :	R^2	0.96941	RPD	4.5997	SEP	0.25942
Bruins 310002:	R^2	0.73627	RPD	1.9275	SEP	0.64582
Bruins 6118 :	R^2	0.92056	RPD	3.5168	SEP	0.35622
Bruins 6175 :	R^2	0.97255	RPD	5.2641	SEP	0.2217
ANN parameters used:	pca	1.00E-08	nodes	3	epochs	1200
Instrument 6110 Robustness:					SEP	0.1797
Instrument 6118 Robustness:					SEP	0.32608
Instrument 6175 Robustness:					SEP	0.25541
Instrument 31002 Robustness:					SEP	0.29908
Bias Corrected Instrument 6118:	R^2	0.92056	RPD	3.5267	SEP	0.35522
Bias Corrected Instrument 310002:	R^2	0.73627	RPD	1.9426	SEP	0.6408

Figure C9. Matlab Script Output. Oil predictions for 'Bruins with MSC pretreatment'

ANN parameters used: pca		1e-008 nodes	3 epochs		1200	
Values below are calculated for MOISTURE						
Bruins 6110 :	R^2	0.99171	RPD	10.983	SEP	0.1976
Bruins 6110 :	R^2	0.99271	RPD	11.539	SEP	0.11199
Bruins 310002:	R^2	0.98907	RPD	9.2837	SEP	0.1392
Bruins 6118 :	R^2	0.99291	RPD	11.733	SEP	0.10997
Bruins 6175 :	R^2	0.99023	RPD	10.094	SEP	0.12619
ANN parameters used:	pca	1.00E-08	nodes	3	epochs	1200
Instrument 6110 Robustness:					SEP	0.15807
Instrument 6118 Robustness:					SEP	0.26092
Instrument 6175 Robustness:					SEP	0.22581
Instrument 31002 Robustness:					SEP	0.56221
Bias Corrected Instrument 6118:	R^2	0.99291	RPD	11.864	SEP	0.10876
Bias Corrected Instrument 310002:	R^2	0.98907	RPD	7.2291	SEP	0.17876

Figure C10. Matlab Script Output. Moisture predictions for 'Bruins with DS standardization'

ANN parameters used: pca	1e-008 nodes		3 epochs		1200	
Values below are calculated for PROTEIN						
Bruins 6110 :	R^2	0.96649	RPD	5.4625	SEP	0.25544
Bruins 6110 :	R^2	0.98348	RPD	7.771	SEP	0.26298
Bruins 310002:	R^2	0.98055	RPD	7.1654	SEP	0.2852
Bruins 6118 :	R^2	0.98556	RPD	8.2808	SEP	0.25052
Bruins 6175 :	R^2	0.97881	RPD	6.8635	SEP	0.28983
ANN parameters used:	pca	1.00E-08	nodes	3	epochs	1200
Instrument 6110 Robustness:					SEP	0.18397
Instrument 6118 Robustness:					SEP	0.19312
Instrument 6175 Robustness:					SEP	0.22917
Instrument 31002 Robustness:					SEP	0.37934
Bias Corrected Instrument 6118:	R^2	0.98556	RPD	7.4347	SEP	0.27903
Bias Corrected Instrument 310002:	R^2	0.98055	RPD	4.4384	SEP	0.46043

Figure C11. Matlab Script Output. Protein predictions for 'Bruins with DS standardization'

ANN parameters used: pca	1e-008 nodes		3 epochs		1200	
Values below are calculated for OIL						
Bruins 6110 :	R^2	0.95726	RPD	4.837	SEP	0.2572
Bruins 6110 :	R^2	0.96562	RPD	5.1875	SEP	0.23003
Bruins 310002:	R^2	0.94881	RPD	4.2881	SEP	0.27827
Bruins 6118 :	R^2	0.97167	RPD	5.5766	SEP	0.21653
Bruins 6175 :	R^2	0.97225	RPD	6.0032	SEP	0.19441
ANN parameters used:	pca	1.00E-08	nodes	3	epochs	1200
Instrument 6110 Robustness:					SEP	0.10686
Instrument 6118 Robustness:					SEP	0.20783
Instrument 6175 Robustness:					SEP	0.22205
Instrument 31002 Robustness:					SEP	0.40933
Bias Corrected Instrument 6118:	R^2	0.97167	RPD	5.936	SEP	0.20342
Bias Corrected Instrument 310002:	R^2	0.94881	RPD	4.2065	SEP	0.28367

Figure C12. Matlab Script Output. Oil predictions for 'Bruins with DS standardization'

ANN parameters used: pca	1e-008 nodes		3 epochs		1200	
Values below are calculated for MOISTURE						
Bruins 6110 :	R^2	0.99171	RPD	10.983	SEP	0.1976
Bruins 6110 :	R^2	0.99271	RPD	11.539	SEP	0.11199
Bruins 310002:	R^2	0.93181	RPD	2.574	SEP	0.50204
Bruins 6118 :	R^2	0.8964	RPD	2.2615	SEP	0.57053
Bruins 6175 :	R^2	0.99023	RPD	10.094	SEP	0.12619
ANN parameters used:	pca	1.00E-08	nodes	3	epochs	1200
Instrument 6110						
Robustness:					SEP	0.18757
Instrument 6118						
Robustness:					SEP	0.50673
Instrument 6175						
Robustness:					SEP	0.32258
Instrument 31002						
Robustness:					SEP	0.25399
Bias Corrected						
Instrument 6118:	R^2	0.8964	RPD	2.2223	SEP	0.5806
Bias Corrected						
Instrument 310002:	R^2	0.93181	RPD	2.7941	SEP	0.4625

Figure C13. Matlab Script Output. Moisture predictions for 'Bruins using PDS model'

ANN parameters used: pca	1e-008 nodes		3 epochs		1200	
Values below are calculated for PROTEIN						
Bruins 6110 :	R^2	0.96649	RPD	5.4625	SEP	0.25544
Bruins 6110 :	R^2	0.98348	RPD	7.771	SEP	0.26298
Bruins 310002:	R^2	0.95865	RPD	2.8648	SEP	0.71335
Bruins 6118 :	R^2	0.89521	RPD	2.5311	SEP	0.81959
Bruins 6175 :	R^2	0.97881	RPD	6.8635	SEP	0.28983
ANN parameters used:	pca	1.00E-08	nodes	3	epochs	1200
Instrument 6110						
Robustness:					SEP	0.35303
Instrument 6118						
Robustness:					SEP	0.35346
Instrument 6175						
Robustness:					SEP	0.51175
Instrument 31002						
Robustness:					SEP	0.54323
Bias Corrected						
Instrument 6118:	R^2	0.89521	RPD	2.3429	SEP	0.88544
Bias Corrected						
Instrument 310002:	R^2	0.95865	RPD	2.1816	SEP	0.93673

Figure C14. Matlab Script Output. Protein predictions for 'Bruins using PDS model'

ANN parameters used: pca	1e-008 nodes		3 epochs		1200	
Values below are calculated for OIL						
Bruins 6110 :	R^2	0.95726	RPD	4.837	SEP	0.2572
Bruins 6110 :	R^2	0.96562	RPD	5.1875	SEP	0.23003
Bruins 310002:	R^2	0.9184	RPD	2.9491	SEP	0.40462
Bruins 6118 :	R^2	0.91102	RPD	2.645	SEP	0.45652
Bruins 6175 :	R^2	0.97225	RPD	6.0032	SEP	0.19441
ANN parameters used:	pca	1.00E-08	nodes	3	epochs	1200
Instrument 6110						
Robustness:					SEP	0.20209
Instrument 6118						
Robustness:					SEP	0.20702
Instrument 6175						
Robustness:					SEP	0.29666
Instrument 31002						
Robustness:					SEP	0.61804
Bias Corrected						
Instrument 6118:	R^2	0.91102	RPD	2.4844	SEP	0.48603
Bias Corrected						
Instrument 310002:	R^2	0.9184	RPD	2.4994	SEP	0.47742

Figure C15. Matlab Script Output. Oil predictions for 'Bruins using PDS model'

ANN parameters used: pca	1e-008 nodes		3 epochs		1200	
Values below are calculated for MOISTURE						
Instrument Validation:	R^2	0.97177	RPD	5.368	SEP	0.21741
Instrument Calibration:	R^2	0.94748	RPD	4.3635	SEP	0.28494
Instrument 0065:	R^2	0.94667	RPD	4.3304	SEP	0.2695
Instrument 3108:	R^2	0.97617	RPD	5.0165	SEP	0.2385
Instrument 553792:	R^2	0.97042	RPD	5.2917	SEP	0.22055
Instrument 1241						
Robustness:	SEP	0.054433				
Instrument 0065						
Robustness:	SEP	0.065396				
Instrument 3108						
Robustness:	SEP	0.10286				
Instrument 553792						
Robustness:	SEP	0.045311				

Figure C16. Matlab Script Output. Moisture predictions for 'FOSS without standardization'

ANN parameters used: pca	1e-008 nodes	3 epochs	1200			
Values below are calculated for PROTEIN						
Instrument Validation:	R^2	0.97899	RPD	6.7964	SEP	0.29269
Instrument Calibration:	R^2	0.97274	RPD	6.0561	SEP	0.22975
Instrument 0065:	R^2	0.98194	RPD	6.9367	SEP	0.28677
Instrument 3108:	R^2	0.9795	RPD	6.8273	SEP	0.29922
Instrument 553792:	R^2	0.96735	RPD	5.2423	SEP	0.37946
Instrument 1241 Robustness:	SEP	0.082813				
Instrument 0065 Robustness:	SEP	0.12697				
Instrument 3108 Robustness:	SEP	0.19482				
Instrument 553792 Robustness:	SEP	0.12154				

Figure C17. Matlab Script Output. Protein predictions for 'FOSS without standardization'

ANN parameters used: pca	1e-008 nodes		3 epochs		1200	
Values below are calculated for OIL						
Instrument Validation:	R^2	0.97199	RPD	5.5165	SEP	0.21156
Instrument Calibration:	R^2	0.94799	RPD	4.3846	SEP	0.28357
Instrument 0065:	R^2	0.94753	RPD	4.3656	SEP	0.26734
Instrument 3108:	R^2	0.97898	RPD	6.2519	SEP	0.19137
Instrument 553792:	R^2	0.97029	RPD	5.7588	SEP	0.20266
Instrument 1241 Robustness:	SEP	0.067197				
Instrument 0065 Robustness:	SEP	0.068466				
Instrument 3108 Robustness:	SEP	0.11061				
Instrument 553792 Robustness:	SEP	0.051017				

Figure C18. Matlab Script Output. Oil predictions for 'FOSS without standardization'

ANN parameters used: pca	1e-008 nodes		3 epochs		1200	
Values below are calculated for MOISTURE						
Instrument Validation:	R^2	0.96588	RPD	5.1712	SEP	0.22569
Instrument Calibration:	R^2	0.96082	RPD	5.0504	SEP	0.24619
Instrument 0065:	R^2	0.92888	RPD	3.7362	SEP	0.31237
Instrument 3108:	R^2	0.96651	RPD	4.9939	SEP	0.23958
Instrument 553792:	R^2	0.84806	RPD	2.4819	SEP	0.47024
Instrument 1241 Robustness:	SEP	0.078128				
Instrument 0065 Robustness:	SEP	0.087475				
Instrument 3108 Robustness:	SEP	0.17124				
Instrument 553792 Robustness:	SEP	0.11443				

Figure C19. Matlab Script Output. Moisture predictions for 'FOSS with FIR standardization'

ANN parameters used: pca	1e-008 nodes	3 epochs	1200			
Values below are calculated for PROTEIN						
Instrument Validation:	R^2	0.96895	RPD	5.5417	SEP	0.35896
Instrument Calibration:	R^2	0.96954	RPD	5.7296	SEP	0.24285
Instrument 0065:	R^2	0.97558	RPD	5.8883	SEP	0.33783
Instrument 3108:	R^2	0.97143	RPD	5.8092	SEP	0.35166
Instrument 553792:	R^2	0.96146	RPD	5.0709	SEP	0.39228
Instrument 1241 Robustness:	SEP	0.085996				
Instrument 0065 Robustness:	SEP	0.12333				
Instrument 3108 Robustness:	SEP	0.307				
Instrument 553792 Robustness:	SEP	0.1061				

Figure C20. Matlab Script Output. Protein predictions for 'FOSS with FIR standardization'

ANN parameters used: pca	1e-008 nodes	3 epochs	1200			
Values below are calculated for OIL						
Instrument Validation:	R^2	0.9811	RPD	5.3925	SEP	0.21643
Instrument Calibration:	R^2	0.96549	RPD	5.3822	SEP	0.23101
Instrument 0065:	R^2	0.95226	RPD	3.2979	SEP	0.35388
Instrument 3108:	R^2	0.97379	RPD	4.8676	SEP	0.2458
Instrument 553792:	R^2	0.97562	RPD	4.346	SEP	0.26854
Instrument 1241 Robustness:	SEP	0.077989				
Instrument 0065 Robustness:	SEP	0.091946				
Instrument 3108 Robustness:	SEP	0.18074				
Instrument 553792 Robustness:	SEP	0.17989				

Figure C20. Matlab Script Output. Oil predictions for 'FOSS with FIR standardization'

ANN parameters used: pca	1e-008 nodes	3 epochs	1200			
Values below are calculated for MOISTURE						
Instrument Validation:	R^2	0.96641	RPD	5.0673	SEP	0.23032
Instrument Calibration:	R^2	0.94861	RPD	4.4112	SEP	0.28186
Instrument 0065:	R^2	0.95133	RPD	4.2324	SEP	0.27575
Instrument 3108:	R^2	0.96583	RPD	4.4978	SEP	0.26601
Instrument 553792:	R^2	0.96187	RPD	4.5433	SEP	0.25688
Instrument 1241 Robustness:	SEP	0.054305				
Instrument 0065 Robustness:	SEP	0.031276				
Instrument 3108 Robustness:	SEP	0.10126				
Instrument 3108 Robustness:	SEP	0.10126				

Figure C21. Matlab Script Output. Moisture predictions for 'FOSS with MSC pretreatment'

ANN parameters used: pca	1e-008 nodes		3 epochs		1200	
Values below are calculated for PROTEIN						
Instrument Validation:	R^2	0.98075	RPD	6.9741	SEP	0.28523
Instrument Calibration:	R^2	0.97242	RPD	6.0218	SEP	0.23106
Instrument 0065:	R^2	0.98154	RPD	7.298	SEP	0.27257
Instrument 3108:	R^2	0.97864	RPD	6.8408	SEP	0.29863
Instrument 553792:	R^2	0.96663	RPD	5.3622	SEP	0.37098
Instrument 1241 Robustness:	SEP	0.11785				
Instrument 0065 Robustness:	SEP	0.12151				
Instrument 3108 Robustness:	SEP	0.34962				
Instrument 3108 Robustness:	SEP	0.34962				

Figure C22. Matlab Script Output. Protein predictions for 'FOSS with MSC pretreatment'

ANN parameters used: pca	1e-008 nodes	3 epochs		1200		
Values below are calculated for OIL						
Instrument Validation:	R^2	0.97199	RPD	5.5165	SEP	0.21156
Instrument Calibration:	R^2	0.94799	RPD	4.3846	SEP	0.28357
Instrument 0065:	R^2	0.95623	RPD	4.5336	SEP	0.25743
Instrument 3108:	R^2	0.96934	RPD	4.9779	SEP	0.24035
Instrument 553792:	R^2	0.96458	RPD	4.8844	SEP	0.23894
Instrument 1241 Robustness:	SEP	0.049903				
Instrument 0065 Robustness:	SEP	0.027578				
Instrument 3108 Robustness:	SEP	0.086488				
Instrument 3108 Robustness:	SEP	0.086488				

Figure C23. Matlab Script Output. Oil predictions for 'FOSS with MSC pretreatment'

WAVELETS type=db6	level=5	Inputs=23	
Values below are calculated for MOISTURE			
Bruins 6110 :	R^2=0.99202	RPD=11.194	SEP=0.19389
Bruins 6110 :	R^2=0.98947	RPD=9.3796	SEP=0.13777
Bruins 310002:	R^2=0.98381	RPD=7.8414	SEP=0.1648
Bruins 6118 :	R^2=0.97854	RPD=6.7532	SEP=0.19106
Bruins 6175 :	R^2=0.98581	RPD=8.3754	SEP=0.15209
ANN parameters used:	pca=1e-008	nodes=3	epochs=1200
Instrument 6110 Robustness:			SEP=0.13458
Instrument 6118 Robustness:			SEP=0.60099
Instrument 6175 Robustness:			SEP=0.13208
Instrument 31002 Robustness:			SEP=0.15652

Figure C24. Matlab Script Output. ANN for Bruins using Wavelets - Moisture

WAVELETS type=db3		level=4	Inputs=29	
Values below are calculated for MOISTURE				
Bruins 6110 :	R^2=0.99233		RPD=11.416	SEP=0.19011
Bruins 6110 :	R^2=0.99268		RPD=10.503	SEP=0.12303
Bruins 310002:	R^2=0.98862		RPD=7.7979	SEP=0.16572
Bruins 6118 :	R^2=0.97411		RPD=5.8517	SEP=0.22049
Bruins 6175 :	R^2=0.99039		RPD=9.895	SEP=0.12873
ANN parameters used:	pca=1e-008		nodes=3	epochs=1200
Instrument 6110				
Robustness:				SEP=0.15848
Instrument 6118				
Robustness:				SEP=0.85293
Instrument 6175				
Robustness:				SEP=0.23119
Instrument 31002				
Robustness:				SEP=0.42843

Figure C25. Matlab Script Output. ANN for Bruins using Wavelets – Moisture

ANN parameters used: pca=1e-008 nodes=3 epochs=1200			
Values below are calculated for PROTEIN			
Bruins 6110 :	R^2=0.9622	RPD=5.1434	SEP=0.27128
Bruins 6110 :	R^2=0.98302	RPD=7.2316	SEP=0.28259
Bruins 310002:	R^2=0.94713	RPD=4.281	SEP=0.47736
Bruins 6118 :	R^2=0.98263	RPD=6.1126	SEP=0.33938
Bruins 6175 :	R^2=0.97584	RPD=6.2904	SEP=0.31623
ANN parameters used:	pca=1e-008	nodes=3	epochs=1200
Instrument 6110			
Robustness:			SEP=0.12852
Instrument 6118			
Robustness:			SEP=0.39784
Instrument 6175			
Robustness:			SEP=0.2038
Instrument 31002			
Robustness:			SEP=0.75294

Figure C25. Matlab Script Output. ANN for Bruins using Wavelets – Protein

WAVELETS type=db3	level=4	Inputs=29	
Values below are calculated for PROTEIN			
Bruins 6110 :	R^2=0.96651	RPD=5.4639	SEP=0.25537
Bruins 6110 :	R^2=0.98333	RPD=7.7424	SEP=0.26395
Bruins 310002:	R^2=0.96439	RPD=5.276	SEP=0.38733
Bruins 6118 :	R^2=0.98077	RPD=6.2938	SEP=0.3296
Bruins 6175 :	R^2=0.97758	RPD=6.6547	SEP=0.29892
ANN parameters used:	pca=1e-008	nodes=3	epochs=1200
Instrument 6110			
Robustness:			SEP=0.080815
Instrument 6118			
Robustness:			SEP=0.35195
Instrument 6175			
Robustness:			SEP=0.14839
Instrument 31002			
Robustness:			SEP=0.34513

Figure C26. Matlab Script Output. ANN for Bruins using Wavelets – Protein

WAVELETS type=db6	level=5	Inputs=23	
Values below are calculated for OIL			
Bruins 6110 :	R^2=0.94553	RPD=4.2847	SEP=0.29035
Bruins 6110 :	R^2=0.96153	RPD=4.858	SEP=0.24563
Bruins 310002:	R^2=0.9475	RPD=4.3152	SEP=0.27652
Bruins 6118 :	R^2=0.97022	RPD=5.4095	SEP=0.22322
Bruins 6175 :	R^2=0.95486	RPD=4.5666	SEP=0.25556
ANN parameters used:	pca=1e-008	nodes=3	epochs=1200
Instrument 6110			
Robustness:			SEP=0.092003
Instrument 6118			
Robustness:			SEP=0.17798
Instrument 6175			
Robustness:			SEP=0.11477
Instrument 31002			
Robustness:			SEP=0.15404

Figure C26. Matlab Script Output. ANN for Bruins using Wavelets – Oil

WAVELETS type=db3	level=4	Inputs=29	
Values below are calculated for OIL			
Bruins 6110 :	R^2=0.95024	RPD=4.4826	SEP=0.27753
Bruins 6110 :	R^2=0.96149	RPD=4.0816	SEP=0.29235
Bruins 310002:	R^2=0.95766	RPD=3.8837	SEP=0.30725
Bruins 6118 :	R^2=0.96555	RPD=4.3675	SEP=0.27647
Bruins 6175 :	R^2=0.96444	RPD=5.0102	SEP=0.23294
ANN parameters used:	pca=1e-008	nodes=3	epochs=1200
Instrument 6110 Robustness:			SEP=0.16172
Instrument 6118 Robustness:			SEP=0.20267
Instrument 6175 Robustness:			SEP=0.17964
Instrument 31002 Robustness:			SEP=0.46258

Figure C27. Matlab Script Output. ANN for Bruins using Wavelets – Oil

APPENDIX D. TABLES WITH RESULTS

TABLE D1. Direct Standardization (DS) results for FOSS and BRUINS using FOSS ANN calibration model

Tested Property	Protein			Oil		
ANN model with DS	R2	RPD	SEP	R2	RPD	SEP
FOSS 1241 master	0.978	6.7	0.292	0.971	5.5	0.211
FOSS 3108	0.974	5.8	0.346	0.974	4.7	0.253
FOSS 553792	0.581	1.1	1.779	0.937	3.8	0.303
FOSS 0065	0.775	2.1	0.981	0.978	6.1	0.195
BRUINS 6110	0.979	6.9	0.269	0.973	5.5	0.221
BRUINS 6118	0.977	6.5	0.287	0.979	6.3	0.194
BRUINS 6175	0.977	6.5	0.286	0.980	6.4	0.190
BRUINS 31002	0.975	6.3	0.296	0.978	6.2	0.196

TABLE D2. Direct Standardization (DS) results for FOSS and BRUINS using PLS calibration model

PROTEIN	Protein					Oil				
PLS - DS	PCs	RPD	SEP	Slope	Bias	PCs	RPD	SEP	Slope	Bias
FOSS 1241 master	12	6.2	0.318	0.979	0.18	11	6.1	0.190	1.010	-0.04
FOSS 3108	12	5.4	0.372	0.965	0.29	11	5.3	0.223	1.003	-0.01
FOSS 553792	12	1.2	1.587	1.115	-0.97	11	1.5	0.775	1.136	-0.57
FOSS 0065	12	2.1	0.995	0.760	2.05	11	7.0	0.174	1.015	-0.06
Bruins 6110	12	6.3	0.296	0.972	0.23	11	6.1	0.201	1.017	-0.07
Bruins 6118	12	6.0	0.311	0.971	0.24	11	6.6	0.186	1.018	-0.07
Bruins 6175	12	6.0	0.311	0.971	0.24	11	7.0	0.175	1.020	-0.08
Bruins 31002	12	6.1	0.308	0.972	0.23	11	7.2	0.169	1.021	-0.09

TABLE D3. Bruins Partial Least Squares (PLS) model using no standardization

BRUINS no standardisation							
PLS model	**BRUINS**	**PROPERTY**	**# PC**	**RPD**	**SEP**	**Slope**	**Bias**
Calibration	Bruins 6110	Moisture	7	5.1	0.426	0.961	0.45
Validation	Bruins 6110	Moisture	7	7.5	0.171	1.021	-0.27
Standardization	Bruins 310002	Moisture	7	5.1	0.249	1.068	-0.88
Standardization2	Bruins 6118	Moisture	7	5.3	0.241	1.029	-0.38
Standardization3	Bruins 6175	Moisture	7	6.1	0.207	0.975	0.32
Calibration	Bruins 310002	Moisture	9	4.1	0.523	1.071	-0.84
Calibration	Bruins 6118	Moisture	12	5.4	0.395	0.936	0.76
Calibration	Bruins 6110	Protein	11	3.8	0.359	0.934	0.56
Validation	Bruins 6110	Protein	11	6.0	0.339	0.976	0.20
Standardization	Bruins 310002	Protein	11	5.2	0.389	1.046	-0.39
Standardization2	Bruins 6118	Protein	11	5.6	0.366	0.997	0.02
Standardization3	Bruins 6175	Protein	11	5.4	0.365	0.948	0.44
Calibration	Bruins 310002	Protein	12	3.1	0.446	1.018	-0.15
Calibration	Bruins 6118	Protein	13	4.5	0.307	0.936	0.54
Calibration	Bruins 6110	Oil	12	3.8	0.324	0.932	0.28
Validation	Bruins 6110	Oil	12	5.6	0.211	1.003	-0.01
Standardization	Bruins 310002	Oil	12	3.9	0.300	0.999	0.00
Standardization2	Bruins 6118	Oil	12	5.3	0.224	1.001	-0.01
Standardization3	Bruins 6175	Oil	12	5.7	0.205	0.964	0.15
Calibration	Bruins 310002	Oil	10	2.9	0.418	0.978	0.09
Calibration	Bruins 6118	Oil	13	3.8	0.327	0.926	0.30

TABLE D4. Bruins Partial Least Squares (PLS) model using FIR standardization

BRUINS Finite Impulse Response standardization							
PLS model	**BRUINS**	**PROPERTY**	**# PC**	**RPD**	**SEP**	**Slope**	**Bias**
Calibration	Bruins 6110	Moisture	12	7.7	0.28	0.983	0.20
Validation	Bruins 6110	Moisture	12	7.7	0.17	0.958	0.54
Standardization	Bruins 310002	Moisture	12	5.0	0.26	0.919	1.05
Standardization2	Bruins 6118	Moisture	12	5.9	0.22	0.953	0.61
Standardization3	Bruins 6175	Moisture	12	7.7	0.16	0.951	0.63
Calibration	Bruins 310002	Moisture	12	4.3	0.5	0.895	1.24
Calibration	Bruins 6118	Moisture	12	7.0	0.31	0.958	0.50
Calibration	Bruins 6110	Protein	12	3.5	0.4	0.92	0.68
Validation	Bruins 6110	Protein	12	4.7	0.43	0.873	1.07
Standardization	Bruins 310002	Protein	12	3.0	0.66	0.793	1.75
Standardization2	Bruins 6118	Protein	12	4.2	0.49	0.888	0.96
Standardization3	Bruins 6175	Protein	12	3.1	0.63	0.788	1.80
Calibration	Bruins 310002	Protein	11	1.5	0.89	0.728	2.30
Calibration	Bruins 6118	Protein	12	3.0	0.45	0.914	0.72
Calibration	Bruins 6110	Oil	12	4.2	0.29	0.945	0.22
Validation	Bruins 6110	Oil	12	5.5	0.22	1.014	-0.06
Standardization	Bruins 310002	Oil	12	2.9	0.41	0.835	0.69
Standardization2	Bruins 6118	Oil	12	5.5	0.22	1.034	-0.15
Standardization3	Bruins 6175	Oil	12	5.4	0.21	1.022	-0.09
Calibration	Bruins 310002	Oil	12	1.5	0.81	0.801	0.83
Calibration	Bruins 6118	Oil	12	3.5	0.36	0.951	0.20

TABLE D5. Bruins Partial Least Squares (PLS) model using MSC pretreatment

BRUINS Multiplicative Scatter Correction							
PLS model	**BRUINS**	**PROPERTY**	**# PC**	**RPD**	**SEP**	**Slope**	**Bias**
Calibration	Bruins 6110	Moisture	6	4.9	0.44	0.960	0.47
Validation	Bruins 6110	Moisture	6	6.1	0.21	1.040	-0.52
Standardization	Bruins 310002	Moisture	6	6.6	0.2	0.918	1.06
Standardization2	Bruins 6118	Moisture	6	6.8	0.19	0.969	0.40
Standardization3	Bruins 6175	Moisture	6	6.9	0.18	0.99	0.13
Calibration	Bruins 310002	Moisture	6	6.5	0.33	0.957	0.51
Calibration	Bruins 6118	Moisture	6	7.1	0.3	0.989	0.13
Calibration	Bruins 6110	Protein	9	2.8	0.49	0.878	1.03
Validation	Bruins 6110	Protein	9	6.4	0.32	0.941	0.50
Standardization	Bruins 310002	Protein	9	3.7	0.55	0.803	1.67
Standardization2	Bruins 6118	Protein	9	4.0	0.51	0.799	1.72
Standardization3	Bruins 6175	Protein	9	2.9	0.68	0.770	1.95
Calibration	Bruins 310002	Protein	12	3.4	0.40	0.907	0.79
Calibration	Bruins 6118	Protein	13	4.2	0.33	0.990	0.08
Calibration	Bruins 6110	Oil	9	3.6	0.34	0.925	0.31
Validation	Bruins 6110	Oil	9	5.2	0.23	0.981	0.08
Standardization	Bruins 310002	Oil	9	4.0	0.3	0.866	0.56
Standardization2	Bruins 6118	Oil	9	5.7	0.21	0.959	0.17
Standardization3	Bruins 6175	Oil	9	5.4	0.21	0.961	0.16
Calibration	Bruins 310002	Oil	9	3.1	0.4	0.894	0.44
Calibration	Bruins 6118	Oil	9	3.6	0.34	0.979	0.08

TABLE D6. Bruins Partial Least Squares (PLS) model using DS standardization

BRUINS Direct Standardization							
PLS	**BRUINS**	**PROPERTY**	**# PC**	**RPD**	**SEP**	**Slope**	**Bias**
Calibration	Bruins 6110	Moisture	7	5.1	0.43	0.961	0.45
Validation	Bruins 6110	Moisture	7	7.5	0.17	1.021	-0.27
Standardization	Bruins 310002	Moisture	7	7.4	0.17	1.022	-0.29
Standardization2	Bruins 6118	Moisture	7	6.8	0.19	1.016	-0.21
Standardization3	Bruins 6175	Moisture	7	6.1	0.21	0.975	0.32
Calibration	Bruins 310002	Moisture	7	3.7	0.58	1.038	-0.45
Calibration	Bruins 6118	Moisture	7	4.4	0.49	0.89	1.31
Calibration	Bruins 6110	Protein	9	2.8	0.49	0.878	1.03
Validation	Bruins 6110	Protein	9	6.4	0.32	0.941	0.50
Standardization	Bruins 310002	Protein	9	6.1	0.33	0.936	0.54
Standardization2	Bruins 6118	Protein	9	6.4	0.32	0.943	0.48
Standardization3	Bruins 6175	Protein	9	3.4	0.57	0.874	1.06
Calibration	Bruins 310002	Protein	11	2.5	0.56	0.905	0.80
Calibration	Bruins 6118	Protein	10	3.3	0.41	0.909	0.77
Calibration	Bruins 6110	Oil	11	3.8	0.32	0.932	0.28
Validation	Bruins 6110	Oil	11	5.66	0.21	1.003	-0.01
Standardization	Bruins 310002	Oil	11	4.5	0.26	0.988	0.04
Standardization2	Bruins 6118	Oil	11	5.9	0.20	1.015	-0.06
Standardization3	Bruins 6175	Oil	11	5.7	0.20	0.964	0.15
Calibration	Bruins 310002	Oil	9	2.7	0.45	0.854	0.61
Calibration	Bruins 6118	Oil	9	3.7	0.34	0.92	0.33

TABLE D7. Bruins Partial Least Squares (PLS) model using PDS standardization

BRUINS Piecewise Direct Standardization							
PLS	**BRUINS**	**PROPERTY**	**# PC**	**RPD**	**SEP**	**Slope**	**Bias**
Calibration	Bruins 6110	Moisture	7	5.1	0.43	0.961	0.45
Validation	Bruins 6110	Moisture	7	7.5	0.17	1.021	-0.27
Standardization	Bruins 310002	Moisture	7	2.4	0.53	0.643	4.65
Standardization2	Bruins 6118	Moisture	7	2.3	0.55	0.634	4.79
Standardization3	Bruins 6175	Moisture	7	6.1	0.21	0.975	0.32
Calibration	Bruins 310002	Moisture	7	2.6	0.83	0.668	3.93
Calibration	Bruins 6118	Moisture	7	2.1	1.01	0.567	5.15
Calibration	Bruins 6110	Protein	9	2.8	0.49	0.878	1.03
Validation	Bruins 6110	Protein	9	6.4	0.32	0.941	0.50
Standardization	Bruins 310002	Protein	9	2.8	0.71	0.685	2.66
Standardization2	Bruins 6118	Protein	9	2.0	1.02	0.606	3.36
Standardization3	Bruins 6175	Protein	9	3.4	0.57	0.874	1.06
Calibration	Bruins 310002	Protein	12	2.3	0.59	0.636	3.08
Calibration	Bruins 6118	Protein	12	1.6	0.85	0.563	3.69
Calibration	Bruins 6110	Oil	8	3.5	0.35	0.919	0.33
Validation	Bruins 6110	Oil	8	5.1	0.23	0.975	0.10
Standardization	Bruins 310002	Oil	8	2.8	0.42	0.715	1.19
Standardization2	Bruins 6118	Oil	8	2.4	0.49	0.639	1.53
Standardization3	Bruins 6175	Oil	8	4.8	0.24	0.962	0.16
Calibration	Bruins 310002	Oil	9	2.2	0.54	0.631	1.53
Calibration	Bruins 6118	Oil	8	2.1	0.59	0.571	1.78

TABLE D8. FOSS Partial Least Squares (PLS) model using no standardization

FOSS no standardization							
PLS	**FOSS**	**PROPERTY**	**# PC**	**RPD**	**SEP**	**Slope**	**Bias**
Calibration	FOSS 1241	Moisture	8	3.5	0.346	0.923	0.32
Validation	FOSS 1241	Moisture	8	5.9	0.196	0.994	0.02
Standardization	FOSS 0065	Moisture	8	4.9	0.236	1.012	-0.05
Standardization2	FOSS 3108	Moisture	8	5.2	0.226	0.991	0.03
Standardization3	FOSS 553792	Moisture	8	2.1	0.551	0.942	0.24
Calibration	FOSS 1241	Protein	13	5.0	0.276	0.961	0.33
Validation	FOSS 1241	Protein	13	6.2	0.319	0.993	0.05
Standardization	FOSS 0065	Protein	13	6.8	0.289	1.014	-0.11
Standardization2	FOSS 3108	Protein	13	6.1	0.331	0.993	0.05
Standardization3	FOSS 553792	Protein	13	3.3	0.588	1.056	-0.47
Calibration	FOSS 1241	Oil	14	4.0	0.309	0.938	0.25
Validation	FOSS 1241	Oil	14	6.6	0.175	0.993	0.02
Standardization	FOSS 0065	Oil	14	5.1	0.226	1.021	-0.08
Standardization2	FOSS 3108	Oil	14	5.7	0.207	1.016	-0.06
Standardization3	FOSS 553792	Oil	14	5.0	0.233	1.012	-0.05

TABLE D9. FOSS Partial Least Squares (PLS) model using FIR standardization

FOSS Finite Impulse Response standardization							
PLS	**FOSS**	**PROPERTY**	**# PC**	**RPD**	**SEP**	**Slope**	**Bias**
Calibration	FOSS 1241	Moisture	6	3.6	0.342	0.924	0.31
Validation	FOSS 1241	Moisture	6	6.6	0.174	0.982	0.07
Standardization	FOSS 0065	Moisture	6	5.8	0.198	0.998	0.01
Standardization2	FOSS 3108	Moisture	6	7.0	0.169	0.990	0.04
Standardization3	FOSS 553792	Moisture	6	2.0	0.569	0.920	0.33
Calibration	FOSS 1241	Protein	15	4.8	0.288	0.957	0.36
Validation	FOSS 1241	Protein	15	6.1	0.325	0.938	0.52
Standardization	FOSS 0065	Protein	15	6.5	0.303	0.935	0.55
Standardization2	FOSS 3108	Protein	15	6.2	0.327	0.936	0.54
Standardization3	FOSS 553792	Protein	15	4.2	0.463	0.988	0.09
Calibration	FOSS 1241	Oil	13	4.4	0.280	0.950	0.21
Validation	FOSS 1241	Oil	13	6.2	0.188	1.030	-0.12
Standardization	FOSS 0065	Oil	13	5.5	0.212	1.038	-0.16
Standardization2	FOSS 3108	Oil	13	6.5	0.184	1.033	-0.14
Standardization3	FOSS 553792	Oil	13	2.2	0.520	0.961	0.16

TABLE D10. FOSS Partial Least Squares (PLS) model using MSC standardization

FOSS Multiplicative Scatter Correction							
PLS	**FOSS**	**PROPERTY**	**# PC**	**RPD**	**SEP**	**Slope**	**Bias**
Calibration	FOSS 1241	Moisture	15	4.0	0.307	0.939	0.25
Validation	FOSS 1241	Moisture	15	6.7	0.172	0.999	0.00
Standardization	FOSS 0065	Moisture	15	5.2	0.220	0.950	0.21
Standardization2	FOSS 3108	Moisture	15	6.6	0.179	0.967	0.13
Standardization3	FOSS 553792	Moisture	15	5.8	0.200	0.963	0.15
Calibration	FOSS 1241	Protein	14	5.1	0.272	0.962	0.32
Validation	FOSS 1241	Protein	14	6.3	0.315	1.006	-0.04
Standardization	FOSS 0065	Protein	14	7.0	0.282	0.946	0.45
Standardization2	FOSS 3108	Protein	14	7.1	0.287	0.952	0.41
Standardization3	FOSS 553792	Protein	14	4.5	0.433	0.979	0.17
Calibration	FOSS 1241	Oil	14	4.0	0.309	0.938	0.25
Validation	FOSS 1241	Oil	14	6.6	0.175	0.993	0.02
Standardization	FOSS 0065	Oil	14	5.1	0.226	0.951	0.20
Standardization2	FOSS 3108	Oil	14	6.1	0.194	0.963	0.15
Standardization3	FOSS 553792	Oil	14	5.2	0.224	0.934	0.27

TABLE D11. FOSS Partial Least Squares (PLS) model using PDS standardization

FOSS Piecewise Direct Standardization							
PLS	**FOSS**	**PROPERTY**	**# PC**	**RPD**	**SEP**	**Slope**	**Bias**
Calibration	FOSS 1241	Moisture	15	4.0	0.307	0.939	0.25
Validation	FOSS 1241	Moisture	15	6.6	0.175	0.993	0.02
Standardization2	FOSS 3108	Moisture	15	1.6	0.708	0.434	2.40
Standardization3	FOSS 553792	Moisture	15	1.4	0.816	0.358	2.71
Calibration	FOSS 1241	Protein	12	4.9	0.281	0.959	0.34
Validation	FOSS 1241	Protein	12	6.2	0.318	0.979	0.18
Standardization	FOSS 0065	Protein	12	0.1	14.880	5.337	-36.13
Standardization2	FOSS 3108	Protein	12	2.0	0.984	0.664	2.85
Standardization3	FOSS 553792	Protein	12	1.7	1.128	0.720	2.37
Calibration	FOSS 1241	Oil	15	4.0	0.307	0.939	0.25
Validation	FOSS 1241	Oil	15	6.7	0.172	0.999	0.01
Standardization	FOSS 0065	Oil	15	0.1	13.950	8.751	-33.6
Standardization2	FOSS 3108	Oil	15	1.6	0.719	0.444	2.36
Standardization3	FOSS 553792	Oil	15	1.4	0.808	0.373	2.65

TABLE D12. FOSS Partial Least Squares (PLS) model using DS standardization

FOSS Direct Standardization							
PLS	**FOSS**	**PROPERTY**	**# PC**	**RPD**	**SEP**	**Slope**	**Bias**
Calibration	FOSS 1241	Moisture	11	3.9	0.317	0.935	0.27
Validation	FOSS 1241	Moisture	11	6.1	0.190	1.010	-0.04
Standardization2	FOSS 3108	Moisture	11	5.3	0.223	1.003	-0.01
Standardization3	FOSS 553792	Moisture	11	1.5	0.775	1.136	-0.57
Calibration	FOSS 1241	Protein	12	4.9	0.281	0.959	0.34
Validation	FOSS 1241	Protein	12	6.2	0.318	0.979	0.18
Standardization2	FOSS 3108	Protein	12	5.4	0.372	0.965	0.29
Standardization3	FOSS 553792	Protein	12	1.2	1.587	1.115	-0.97
Calibration	FOSS 1241	Oil	11	3.9	0.317	0.935	0.27
Validation	FOSS 1241	Oil	11	6.1	0.190	1.010	-0.04
Standardization2	FOSS 3108	Oil	11	5.3	0.223	1.003	-0.01
Standardization3	FOSS 553792	Oil	11	1.5	0.775	1.136	-0.57

ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who supported me while I completed my dissertation. I am deeply indebted to my major professor, Dr. Charles R. Hurburgh, Jr. as well as Glen Rippke and Darren Jarboe of the Grain Quality Initiative, whose help, stimulating discussions, and encouragement helped me during the researching and writing of this dissertation. The members of my POS committee also deserve my thanks for their support of this project and their excellent feedback. The graduate students, undergraduate students, and other student workers in the Grain Quality Lab were also important to the success of my research. I also want to thank Gretchen Mosher for her help with the intricate rules of English grammar.